UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE FILOSOFÍA Y LETRAS

DEPARTAMENTO DE FILOLOGÍA INGLESA

Programa de Doctorado: Lingüística Aplicada

TESIS DOCTORAL

Assessment for Learning in Primary CLIL Classrooms and its Co-Construction in Classroom Discourse

Irene Pascual Peña

Directora: Dr. Ana Llinares García

Madrid, 2017

Assessment for Learning in Primary CLIL Classrooms and its Co-Construction in Classroom Discourse

Abstract

The present dissertation addresses the role of *Assessment for Learning* (AfL) in primary *Content and Language Integrated Learning* (CLIL) classrooms in the community of Madrid. Black and Wiliam (1998a: 7-8) define AfL as "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged". The main purpose of the study is to characterize how discourse is jointly co-constructed by teachers and students in whole-class interactions in the CLIL/bilingual education programmes under analysis, and to explain how classroom discourse is aligned (or not) with an AfL pedagogy, and its effects on students' language and content use. This study aims to conduct a thorough and detailed analysis of how AfL is constructed in primary CLIL classrooms through teacher-student interactions across a variety of subjects. This research is innovative, as AfL theory has only been applied to CLIL classrooms in one previous study on motivation and assessment for learning using the same contexts studied here (Basse 2016); and hence more research in this area, and on classroom discourse in particular, is clearly needed (Nikula, Dalton-Puffer & Llinares 2013). It is necessary to study AfL in CLIL classrooms not only as a form of assessment, but also to find out whether AfL discourse can have positive effects on the quality of the interaction. This may have important implications for CLIL pedagogy as high-quality interaction in CLIL classrooms has been seen to positively affect students' content and language engagement (Dalton-Puffer 2009; Nikula, Dalton-Puffer & Llinares 2013).

The data used for this research is part of a bigger corpus collected in the academic year 2010-2011. The corpus consists of 500,000 words over a total of 80 class sessions in 5 bilingual primary schools. For the purpose of this dissertation, 44 sessions out of the 80 total were selected from 4 different schools, which makes an approximate total of 300,000 words. The four schools in which data were collected were grouped in two categories: *Non-AfL schools* and *AfL schools*. AfL schools are those in which Assessment for Learning is implemented and teachers have been trained for it, whereas Non-AfL schools are those in which Assessment for Learning is not specifically implemented. The data used for this study were collected in the third cycle of Primary Education: 5th and 6th year, in which students are, correspondingly, 10-11 and 11-12 years old. Each teacher in each of the schools was recorded in two different subjects. All the

data were analysed in different ways to control the various variables in terms of episodes, teachers' questions, students' responses, students' initiations and teachers' feedback.

The results show significant differences between the classes in which AfL was implemented and the non-AfL ones when IRF patterns and episodes were analysed. Findings also indicate that certain types of episodes (*stating objectives for the lesson*, *explaining marks* and *self-/peer-assessment*), teachers' questions (questions for opinions, for reasons, meta-cognitive questions and meta-questions) and teachers' feedback (expansion, re-route or meta-feedback) align with an AfL pedagogy, as they all aim to explore students' thinking and learning processes and to encourage them to reflect on those processes. These discourse features which align with an AfL methodology have a positive impact on students' participation and contributions, thus making them active and engaged in interaction. These features also contribute to create high-quality interaction, in which students are provided with learning opportunities, both content and language learning. The analysis also reveals that the type of subject taught has an effect on the types of question and feedback used by the teacher. This is one of the first investigations which contributes to both CLIL and AfL research and practice adopting a discourse perspective. In the light of this research, it is concluded that implementing AfL in CLIL classrooms can have important benefits for the type of interaction that goes on inside the black box.

Key words: Assessment for Learning (AfL), Content and Language Integrated Learning (CLIL), classroom discourse and interaction, IRF patterns.

Resumen

La presente tesis aborda el papel de la Evaluación para el Aprendizaje (AFL) en las aulas de Aprendizaje Integrado de Contenidos y Lenguas Extranjeras (AICLE) en la Comunidad de Madrid. Black y Wiliam (1998a: 7-8) definen la AfL como "todas aquellas actividades realizadas por profesores y / o por sus estudiantes, que proveen información para ser usada como retroalimentación para modificar las actividades de enseñanza y aprendizaje en las que están comprometidos". El propósito principal del estudio es caracterizar cómo el discurso es co-construido conjuntamente por los profesores y los estudiantes en la interacción de clase en los programas de AICLE / educación bilingüe bajo análisis y para explicar cómo el discurso del aula está alineado (o no) con una pedagogía AfL, así como sus efectos en el uso del idioma y del contenido de los estudiantes. Este estudio tiene como objetivo realizar un análisis minucioso y detallado de cómo AfL se construye en las aulas AICLE de primaria a través de interacciones profesor-alumno a través de una variedad de asignaturas. Esta investigación es innovadora, ya que la teoría AFL sólo se ha aplicado a aulas AICLE en un estudio previo sobre motivación y evaluación para el aprendizaje utilizando los mismos contextos estudiados aquí (Basse 2016) y, por lo tanto, más investigación en esta área, y en el discurso del aula en particular, es claramente necesario (Nikula, Dalton-Puffer & Llinares 2013). Es necesario estudiar AfL en las aulas CLIL no sólo como una forma de evaluación, sino también para averiguar si el discurso AfL puede tener efectos positivos sobre la calidad de la interacción. Esto puede tener implicaciones importantes para la pedagogía AICLE, ya que se ha visto que la interacción de alta calidad en estas aulas afecta positivamente al compromiso con el contenido y con la lengua extranjera por parte de los estudiantes (Dalton-Puffer 2009; Nikula, Dalton-Puffer & Llinares 2013).

Los datos utilizados para esta investigación forman parte de un corpus más grande recogido en el curso académico 2010-2011. El corpus consiste en 500.000 palabras en un total de 80 sesiones de clase en 5 escuelas primarias bilingües. Para el propósito de esta disertación, 44 sesiones de los 80 totales fueron seleccionadas de 4 colegios diferentes, lo que hace un total aproximado de 300.000 palabras. Los cuatro colegios en los que se recolectaron los datos se agruparon en dos categorías: Colegios no AfL y Colegios AfL. Éstos últimos son aquellos en los que se implementa la evaluación para el aprendizaje y se ha formado a los maestros para ello, mientras que los colegios no AfL son aquellos en los que no se implementa específicamente la evaluación para el aprendizaje. Los datos utilizados para este estudio fueron

recogidos en el tercer ciclo de Educación Primaria: 5º y 6º año, en el que los alumnos tienen, respectivamente, 10-11 y 11-12 años. Cada maestro en cada una de las escuelas fue grabado en dos asignaturas diferentes. Todos los datos fueron analizados de diferentes maneras para controlar las distintas variables en términos de episodios, preguntas de los profesores, respuestas de los estudiantes, iniciaciones de los estudiantes y el tercer turno (*feedback*) de los profesores.

Los resultados muestran diferencias significativas entre las clases en las que se implementó AfL y las no AfL cuando se analizaron los patrones de IRF (Iniciación-Respuesta-*Feedback*) y los episodios. Los resultados también indican que ciertos tipos de episodios (indicar objetivos de la lección, explicar notas y la autoevaluación y co-evaluación), ciertos tipos de preguntas de los profesores (preguntas de opinión, de razones, preguntas meta-cognitivas y meta-preguntas), y ciertos tipos de *feedback* (expansión, reorientación o *meta-feedback*) se alinean con una pedagogía del AfL, ya que tienen como objetivo explorar los procesos de pensamiento y aprendizaje de los estudiantes y animarlos a reflexionar sobre dichos procesos. Estos rasgos discursivos que se alinean con una metodología AfL tienen un impacto positivo en la participación y contribución de los estudiantes, haciendo que éstos estén activos y comprometidos en la interacción. Estas características también contribuyen a crear una interacción de alta calidad, en la que los estudiantes reciben oportunidades de aprendizaje, tanto de contenido como de la lengua extranjera. El análisis también revela que el tipo de materia enseñada tiene un efecto sobre los tipos de pregunta y *feedback* utilizados por el profesor. Ésta es una de las primeras investigaciones que contribuye tanto a la investigación como a la práctica de AICLE y AFL adoptando una perspectiva discursiva. A la luz de esta investigación, se concluye que la implementación de AfL en las aulas CLIL puede tener importantes beneficios para el tipo de interacción que se desarrolla dentro del aula.

Palabras clave: Evaluación para el Aprendizaje (AfL), Aprendizaje Integrado de Contenido y Lenguas Extranjeras (AICLE), discurso e interacción en el aula, patrón IRF.

Table of contents

List of Tables

List of Figures

Agradecimientos

En primer lugar, mi más sincero agradecimiento va para mi directora de tesis, Dra. Ana Llinares García. Hace ya mucho tiempo que comencé a trabajar con ella, desde que estudiaba la carrera. Durante todos estos años, siempre me ha apoyado y me ha brindado excelentes oportunidades académicas. Huelga decir que esta tesis es tan mía como suya. Esta tesis nunca habría visto la luz sin su esfuerzo, sin su trabajo, sin su brillantez profesional, sin su apoyo… sin ella, en definitiva. Siempre te estaré agradecida por ser una fuente constante de trabajo, guía e inspiración.

Gracias a la Universidad Autónoma de Madrid, que hizo posible que empezara el doctorado y siguiera con él durante todos estos años a través de la beca de Inicio de Estudios al Tercer Ciclo y la beca de Formación de Personal Investigador (FPI). Este tipo de becas son imprescindibles para ayudar a la investigación en nuestro país, y estoy profundamente agradecida por haber tenido el honor de disfrutar de ellas.

Quiero agradecer también su ayuda al Departamento de Filología Inglesa. Han sido muchos los miembros de este Departamento que me han ayudado, que han mostrado su interés en mi trabajo y que siempre me han apoyado y alentado a lo largo de tantos años, ya desde que era estudiante de la carrera. Dentro de todos estos miembros, es imprescindible que haga mención especial a la Dra. Rachel Whittaker, el Dr. Tom Morton y el Dr. Mick O'Donnell. Rachel Whittaker, porque siempre me ha tratado y ayudado como si fuera su propia doctoranda. Ha sido muy importante su apoyo y todo lo que he aprendido de ella. Tom Morton, porque desde que llegó al equipo siempre ha sido un espejo en el que mirarse, un ejemplo de cómo hacer las cosas bien. Mick O'Donnell, porque diseñó el maravilloso UAM Corpus Tool, el cual he utilizado para todo el análisis de esta tesis. Además, sin su ayuda con esta herramienta no habría podido hacer un análisis tan complejo. Mi más sincero agradecimiento por los ratos que has estado conmigo ayudándome para que todo saliera bien.

Siempre estaré en deuda con el King's College London, por aceptarme como investigadora para hacer mi estancia allí, y con el Dr. Constant Leung, quien me dirigió mientras estaba allí. Su ayuda incansable mejoró notablemente el presente trabajo en aspectos clave de metodología. Gracias por tu brillante trabajo y también, en el plano personal, por la manera en que me acogiste y estuviste pendiente de mí.

Por supuesto, cómo no agradecer a mi eterna compañera, Rachel Basse. Desde que empezamos a recoger juntas el corpus en 2010, han sido muchas cosas vividas juntas, tanto profesionales como personales. Ha sido un camino duro y largo, pero tenerte a mi lado durante todo el viaje ha sido alentador. Sé que sin ti todo habría sido mucho más difícil. Mil gracias por mil cosas.

Durante este largo camino, he tenido más compañeras de viaje. Algunas que lo iniciaron conmigo y no lo acabaron, otras que están a punto de hacerlo. Amanda, gracias por muchas cosas que hemos compartido. Ánimo, que ya casi estás al final del camino también.

En la fase final de este trabajo, en el que las prisas mandaban, la ayuda de Thomas ha sido determinante. Gracias por todo tu esfuerzo para intentar mejorar este trabajo. Te agradezco de veras el trabajo tan rápido y tan bien hecho. Eres un gran profesional.

Mi más sincero agradecimiento a todos los centros, profesores y alumnos que quisieron participar en este proyecto. Gracias a ellos ha sido todo posible. Gracias por contribuir con la investigación y por las ganas de querer hacer las cosas mejor.

Mi familia y mis amigos han sido dos pilares fundamentales para no caerme durante el proceso. Simplemente con tenerles cerca ha bastado. Gracias a todos.

Cómo no, gracias a mis padres. No sólo por el infinito apoyo que me han dado durante esta etapa, sino por todo lo que me han dado en la vida. Por todo lo que me han aguantado. Por todo lo que me han apoyado. Por todo lo que me han enseñado. Siempre ahí. Siempre dispuestos. Eternamente, gracias.

Por último, gracias a Chema, por ser mi compañero, mi amigo, mi todo. He alcanzado la meta contigo. Dudo que hubiera sido capaz de hacerlo sin ti. Simplemente, gracias por darme vida.

Presentación y conclusiones de la tesis

1 PRESENTACIÓN

1.1 Finalidad y alcance del estudio

La presente tesis aborda el papel de la Evaluación para el Aprendizaje (AfL) o Evaluación Formativa en clases de Aprendizaje Integrado de Contenido y Lenguas (AICLE) en la Comunidad de Madrid. El propósito principal del estudio es caracterizar cómo el discurso es co-construido conjuntamente por los profesores y los estudiantes en interacciones de clase en los programas AICLE / educación bilingüe bajo análisis y explicar cómo el discurso del aula está alineado (o no) con una pedagogía AfL, y sus efectos en el uso de la lengua extranjera y del contenido por parte del alumnado.

El Aprendizaje Integrado de Contenido y Lengua (AICLE) se refiere a "cualquier situación educativa en la que se utiliza una lengua adicional y por lo tanto no la lengua más utilizada en el ambiente para la enseñanza y el aprendizaje de materias distintas del propio idioma" (Wolff 2007: 16). AICLE ha sido apoyado durante casi tres décadas por muchos proyectos europeos y ha sido concebido para crear ciudadanos multilingües en una Europa multilingüe (Whittaker y Llinares 2009; Dalton-Puffer 2011). Este método de enseñanza que, como dice el nombre, consiste en la enseñanza integrada de materias de contenido y una lengua extranjera, proporciona a los estudiantes tiempo adicional de exposición a dicha lengua extranjera y práctica en ella (Dale & Tanner 2012), pero también implica un enfoque integrado para aprender la lengua extranjera y el contenido (por ejemplo, Llinares 2015). Muchos países europeos diferentes (más del 80%) están aplicando el AICLE de una manera u otra (Dale y Tanner 2012). España es uno de los países en los que AICLE ha sido implementado muy rápidamente (Llinares y Dafouz 2010). En Madrid, en la enseñanza primaria, se están llevando a cabo dos proyectos AICLE: el proyecto del Ministerio de Educación / British Council y el proyecto de la Comunidad de Madrid (Llinares y Dafouz 2010). En esta tesis, los datos provienen de escuelas que participan en cada uno de estos dos proyectos, siendo el enfoque del estudio la comparación entre los que implementan la Evaluación para el Aprendizaje y los que no lo hacen.

Como resultado de la rápida implementación de AICLE en Europa, su investigación sobre ha sido extensa, tanto a nivel empírico como teórico (Dalton-Puffer 2011). Muchos aspectos han sido investigados a nivel empírico (ver Dalton-Puffer 2011 para una revisión), pero parece que la evaluación es una de las áreas en la investigación de AICLE que apenas se ha explorado, especialmente la Evaluación para el Aprendizaje (pero ver Basse 2016; Llinares , Morton y Whittaker 2012, y Pascual y Basse 2017).

La Evaluación para el Aprendizaje tuvo su punto de inflexión con la revisión de Black y Wiliam (1998a) de la literatura sobre este tipo de evaluación, en la que los autores resaltaron los beneficios de este tipo de metodología, a la que siguieron otros estudios e incluso algunos cambios políticos. Sin embargo, los estudios revisados en la extensa revisión de la literatura de Black y William se centraron principalmente en aulas donde el contenido se estudia en la lengua materna (Rea-Dickins 2008). Hay excepciones, como los estudios de Leung y Mohan (2004), Rea-Dickins (2001), Rea-Dickins y Gardner (2000), y Edelenbos y Kubanek-German (2004), basado en aulas primarias con el inglés como lengua adicional (EAL), el inglés como segunda lengua (ESL), y el inglés como lengua extranjera (EFL). AfL en las aulas AICLE sigue siendo un tema poco investigado al que esta tesis pretende contribuir.

Este estudio tiene como objetivo realizar un análisis minucioso y detallado de cómo AfL se construye en aulas AICLE de primaria a través de interacciones profesor-alumno en una variedad de asignaturas. Esta investigación es innovadora, ya que la teoría AfL sólo se ha aplicado a aulas AICLE en un estudio previo sobre motivación y Evaluación para el Aprendizaje utilizando los mismos contextos estudiados aquí (Basse 2016). Por lo tanto, es necesaria una mayor investigación en esta área, y en el discurso de clase en particular (Nikula, Dalton-Puffer y Llinares 2013). Es necesario estudiar AfL en las aulas AICLE no sólo como una forma de evaluación, sino también para averiguar si el discurso AfL puede tener efectos positivos sobre la calidad de la interacción. Esto puede tener implicaciones importantes para la pedagogía AICLE ya que se ha visto que la interacción de alta calidad en estas aulas afecta positivamente a la implicación con el contenido y con la lengua de los estudiantes (Dalton-Puffer 2009). Por lo tanto, uno de los objetivos de esta tesis es crear un modelo para caracterizar el discurso AfL, que podría ser utilizado como base para cursos de formación de profesores y la creación de materiales para la implementación de AfL en las aulas AICLE.

La teoría de la Evaluación para el Aprendizaje afirma que la enseñanza, el aprendizaje y la evaluación son parte del mismo proceso (Black y Wiliam 1998a, b). Es una teoría que está estrechamente relacionada con la teoría sociocultural, en la que se considera que la interacción es fundamental para el aprendizaje (Vygotsky 1978, Van Lier 1996, Lantolf y Thorne 2006).

Como dicen Leung y Mohan (2004: 336), "gran parte del trabajo formativo se realiza interactivamente a través de la interacción entre profesores y estudiantes". Si la interacción es esencial para el aprendizaje, y si el aprendizaje, la enseñanza y la evaluación son la misma entidad, se deduce que la interacción es también esencial para la Evaluación para el Aprendizaje (Black y Wiliam 1998b). En la misma línea, Leung (2004: 29) expresa la necesidad de prestar atención a la interacción en el aula para estudiar y entender la Evaluación para el Aprendizaje: "la naturaleza socialmente co-construida de la evaluación formativa hace, por tanto, necesario asistir a la interacción y el discurso del aula como un lugar clave para la investigación empírica" (véase también Black y Wiliam 1998a, Edelenbos y Kubanek-German 2004). A pesar de la importancia probada de la interacción en la Evaluación para el Aprendizaje, existen muy pocos estudios que analicen la interacción real en relación con AfL (Leung y Mohan, 2004; Harlen y Winter 2004; Rea-Dickins 2001; Gardner y Rea-Dickins 2002; Anderson et al. 2007; Ruiz-Primo y Furtak 2006, 2007). En este sentido, esta tesis contribuiría a esta área en la investigación AfL en un nuevo contexto (es decir, AICLE).

En resumen, al analizar la interacción en el aula y relacionarla con la teoría de la Evaluación para el Aprendizaje en las aulas AICLE, esta tesis pretende llenar dos vacíos importantes de investigación: uno, caracterizar el discurso de AfL en AICLE y dos, explorar los efectos de la interacción AfL sobre el uso y el aprendizaje de contenido y lengua por parte del alumnado. A través de la implementación del modelo AfL para las aulas AICLE propuesto en esta tesis, los profesores AICLE pueden interactuar con los malentendidos de contenido o lingüísticos integrando los dos aspectos o prestando especial atención a uno u otro. El *feedback* centrado en el contenido también puede ser beneficioso para la competencia lingüística de los estudiantes, y el *feedback* centrado en la lengua también puede ser muy valioso para que los estudiantes alcancen metas de aprendizaje de contenido (Llinares et al. 2012).

1.2 AICLE

El AICLE se ha definido a menudo tanto como un tipo de programa como como un tipo de metodología. Esta dicotomía se refleja en los niveles macro y micro de su investigación (Dalton-Puffer y Smit 2007, véase también Leung 2005a para la educación bilingüe en general): el nivel micro se centra en los participantes, resultados y procesos en las implementaciones metodológicas; el nivel macro se refiere a las características de los programas AICLE. Dalton-Puffer y Smit (2007) enfatizan la importancia del nivel micro, ya que sin él el nivel macro no existiría. Esta disertación está enmarcada en el nivel micro, y contribuirá a este creciente flujo

de investigación en AICLE (ver, por ejemplo, Llinares y Whittaker 2009; Dalton-Puffer 2007; Dalton-Puffer y Nikula 2006).

Se ha afirmado que uno de los aspectos que distinguen a AICLE de otros tipos de educación bilingüe (como la inmersión o la instrucción basada en contenido) es el hecho de que la lengua de instrucción es una lengua extranjera y no una segunda lengua. Además, el hecho de que este idioma extranjero sea en su mayoría el inglés (como en el presente estudio) llevó a Dalton-Puffer et al. (2010) a acuñar el término CEIL (Aprendizaje Integrado de Contenido e Inglés - *Content and English Integrated Learning*-). Otro aspecto distintivo es, como indica el propio término, la integración de los objetivos lingüísticos y de contenido (Dalton-Puffer y Smit 2007). Sin embargo, según Dalton-Puffer y Smit (2007), en Europa la mayoría de los currículos están basados en el contenido, lo que significa que los aspectos lingüísticos a menudo se dejan de lado. Esto podría explicarse por el hecho de que los gobiernos europeos no invierten suficientes recursos en la formación de profesores y en la implementación de AICLE (Dalton-Puffer 2011). En España, los programas previos de formación AICLE son prácticamente inexistentes y los programas de formación en el puesto de trabajo que aún existen son escasos (Fernández Fontecha 2009) y rara vez abordan el papel de la lengua en la enseñanza del contenido.

Los objetivos de aprendizaje de la lengua no sólo deben permanecer visibles en los planes de estudio AICLE, sino que deben ser explícitos tanto para los profesores como para los estudiantes (Llinares, Morton y Whittaker, 2012). Aunque la lengua es un componente clave en cualquier tipo de evaluación, AfL adquiere especial relevancia para AICLE porque es planificado y reactivo (Llinares et al. 2012). Cuando se están planificando los objetivos de aprendizaje, no sólo deben tenerse en cuenta conceptos y habilidades, sino también el lenguaje que va de la mano con ellos (véase también Coyle et al. 2010). Por otra parte, los profesores AICLE también deben ser sensibles a las necesidades de sus estudiantes y, en consecuencia, necesitan reaccionar y ajustar su enseñanza, centrándose en el lenguaje necesario para adquirir diferentes objetivos de contenido. La taxonomía presentada en este estudio explica posibles diferencias entre la atención al lenguaje o la atención al contenido. Aunque se puede argumentar que el contenido y el lenguaje no pueden separarse el uno del otro (Halliday 1978, Widdowson 1978, Coyle et al. 2010), esta distinción se ha hecho para destacar la dualidad de AICLE (contenido y lengua). Precisamente de esta dualidad surge uno de sus principales desafíos: aprender una lengua extranjera y usar esa lengua para comprender, expresar y aprender adecuadamente el contenido en escenarios de enseñanza / aprendizaje donde los niveles cognitivos y lingüísticos de los estudiantes pueden variar sustancialmente (Coyle et al. 2010).

1.3 El marco teórico: Evaluación para el Aprendizaje y el discurso en el aula

Las perspectivas teóricas aplicadas en esta tesis son dos: Evaluación para el Aprendizaje y discurso en el aula.

Ha habido mucha discusión reciente sobre la evaluación por parte de los profesores en el aula. Recientemente, se ha argumentado que la evaluación debe integrarse con la enseñanza, ser formativa y para el aprendizaje (Leung y Mohan 2004). Este tipo de evaluación se ha visto a veces como una alternativa más deseable a la evaluación sumativa, mientras que otros la ven como una práctica complementaria (Leung 2004). La Evaluación para el Aprendizaje, también llamada evaluación formativa o evaluación de profesores en el aula, contrasta con la evaluación sumativa o con pruebas formales, ya que ésta última se centra en el producto del aprendizaje (Spolsky 1992) mientras que la primera se centra en el proceso de aprendizaje (Leung y Mohan 2004, Black y William 1998a, Rea-Dickins 2001). La Evaluación para el Aprendizaje ha sido definida como una forma de evaluación cuyo objetivo final es mejorar las prácticas de enseñanza y los procesos de aprendizaje y los resultados (ver, por ejemplo, Shohamy 1992, Harlen 2005, Black y William 1998a, b). Como señalan algunos investigadores (Leung y Mohan 2004), es necesario examinar en profundidad la evaluación formativa de los profesores, tal como se ha hecho con la evaluación estandarizada, si queremos entender cómo este tipo de evaluación se logra realmente en la interacción en el aula, y si queremos apropiarnos tanto de la teoría como de los métodos de investigación "en el estudio de este aspecto altamente complejo y dinámico de la interfaz enseñanza-aprendizaje" (Leung y Mohan 2004: 338). Si la evaluación formativa se realiza a través de la interacción maestro-alumno, la evaluación de los estudiantes necesariamente tiene que ser re-conceptualizada como discurso: "la evaluación formativa tiene que tomar en cuenta la naturaleza interactiva y contingente del desempeño estudiantil en el aula, que es dinámica y coproducida con el docente y otros "(Leung 2004: 22).

La Evaluación para el Aprendizaje debe integrarse en las actividades cotidianas de enseñanza y aprendizaje y se supone que los profesores pueden responder a las necesidades contingentes de los estudiantes en el proceso de aprendizaje (Leung 2004). Es por eso que la evaluación formativa no puede acomodarse fácilmente dentro de un conjunto de criterios pre-especificados (Leung 2004). La QCA (*Qualifications and Curriculum Authority*, una agencia gubernamental cuasi oficial en Inglaterra) presenta los principios de la Evaluación para el Aprendizaje de la siguiente manera:

La Evaluación para el Aprendizaje ocurre todo el tiempo en el aula. [...] un alumno necesita saber dónde está y entender no sólo dónde quiere estar, sino también cómo "cubrir el hueco". Esto implica tanto al profesor como al alumno en un proceso de reflexión continua y revisión

sobre el progreso. Cuando los profesores y sus compañeros proporcionan *feedback* de calidad, los alumnos están facultados para tomar las medidas apropiadas. Los profesores ajustan sus planes en respuesta a la evaluación formativa (en Leung 2004: 22).

Como sugiere Clarke (1998: 117), "los exámenes formales, aunque útiles para algunos propósitos, no parecen elevar en sí los estándares. Las estrategias formativas de evaluación sí pueden hacerlo". De la misma manera, el trabajo de Black y Wiliam (1998a: 3) demostró que "las innovaciones que incluyen el fortalecimiento de la práctica de la evaluación formativa producen ganancias de aprendizaje significativas y a menudo sustanciales. Estos estudios varían en edades (desde niños de cinco años hasta graduados universitarios), en materias y en países ".

En AfL, los análisis de interacción en el aula son importantes principalmente por dos razones: en primer lugar, porque es en la interacción donde se lleva a cabo la Evaluación para el Aprendizaje (Leung 2004; Leung y Mohan 2004; Rea-Dickins 2001; Black y Wiliam 1998b); y segundo, porque si llegamos a saber cómo funciona este tipo de evaluación, su implementación podría ser más efectiva. Si comprendemos que el aprendizaje es co-construido conjuntamente por los individuos (maestro y estudiantes en el caso de un contexto educativo) (Vygotsky 1978, Hammond y Gibbons 2005), y que uno de los aspectos cruciales para el aprendizaje es la interacción (Van Lier 1988, 1996; Mortimer y Scott 2003; Barnes 1975; Hall y Walsh 2002; Gibbons 2003; Vygotsky 1978; Lantolf y Thorne 2006; Llinares y Whittaker 2009), entonces la interacción es también fundamental para AfL, ya que cualquier cosa que pueda mejorar el aprendizaje es importante para esta pedagogía (Black y Wiliam 1998b, Leung y Mohan en prensa, Leung 2004, Rea-Dickins 2001). Como señalan Leung y Mohan (2004: 336), "gran parte del trabajo formativo se realiza interactivamente a través de la interacción entre docentes y estudiantes". Como la evaluación y el aprendizaje son interactivos y socialmente construidos, el discurso y la interacción en el aula reflejarán y construirán una evaluación formativa. Como consecuencia de esta estrecha relación, se podría concluir que la interacción y el discurso en el aula necesitan una investigación más profunda para comprender la evaluación formativa y su naturaleza co-construida socialmente (Leung 2004: 29). La importancia del discurso se destaca no sólo en AfL, sino también en la teoría sociocultural y otros enfoques sobre la adquisición de una segunda lengua en el aula. Una de las razones por las que esta tesis puede ser una aportación valiosa tanto para la investigación de AfL como de AICLE es su análisis detallado del discurso de clase en las aulas AICLE de primaria y su relación con la pedagogía AfL, siempre orientada a mejorar el aprendizaje.

Uno de los patrones de discurso del aula que ha sido más ampliamente estudiado en relación con las oportunidades que ofrece para el uso del lenguaje y el aprendizaje es el intercambio de

Iniciación-Respuesta-Feedback (IRF) (Sinclair y Coulthard 1975). Este patrón ha sido ampliamente criticado porque los maestros hacen la mayor parte de la conversación y los estudiantes no pueden perseguir sus propios temas o ideas (Barnes, 1975, 1993; Mortimer y Scott 2003). Por el contrario, en otros estudios se ha dicho que este patrón es eficaz para ciertos propósitos, especialmente para comprobar la comprensión de los estudiantes y guiar su aprendizaje (Mercer 1992; Seedhouse 1997; Nassaji y Wells 2000). El presente estudio explorará el papel de los patrones IRF en clases AICLE donde se implementa AfL y otras en las que no se implementa.

Dentro del intercambio IRF, las preguntas de los profesores han sido ampliamente estudiadas desde perspectivas diferentes y en una variedad de tipos de clases. Estas perspectivas incluyen la dicotomía formal abierta / cerrada (Barnes 1969; Musumeci 1996), la dicotomía funcional demostración / referencial (Long y Sato 1983, Romero y Llinares 2001), la perspectiva del contenido (Dalton-Puffer 2007; Llinares y Pascual 2014) y el enfoque de la demanda cognitiva (Bloom et al., 1956; Redfield y Rousseau 1981). La formulación de preguntas por parte de los profesores es la clave para una enseñanza exitosa (Wragg y Brown 2001) y, por lo tanto, una de las características más importantes de AfL, ya que está comprometida con el éxito de la enseñanza y el aprendizaje (Black et al. 2004; Black et al. 2003; Wiliam et al. 2004; Black y Wiliam 1998b). Varios estudios han afirmado que las preguntas de los profesores deben fomentar la discusión, revelar la comprensión de los estudiantes y requerir respuestas largas (Harrison y Howard 2009). En esta tesis, con ayuda de una taxonomía específica, se analizan tanto las preguntas de los docentes como las correspondientes respuestas de los alumnos, ya que las aportaciones de éstos también forman parte del discurso de clase y son cruciales para que los docentes guíen su aprendizaje (Van Lier 1996).

La tercera parte del intercambio IRF (*feedback*) también ha sido objeto de mucha investigación debido a su importancia para una interacción de calidad (Lee 2007; Lyster 2007; Alexander 2004; Nassaji y Wells 2000). Algunos investigadores distinguen entre IRF y el IRE (Iniciación-Respuesta-Evaluación), dependiendo de si el profesor ofrece *feedback* evaluativo o no-evaluativo (Mehan 1979; Van Lier 1996; Hall y Walsh 2002; Wells 1993). Una característica básica de AfL es que el *feedback* debe ser útil para guiar y mejorar, y debe apuntar hacia el progreso del aprendizaje (Leung 2007; Harrison y Howard 2009; Leahy et al. 2005). Una de las formas en que esto se puede lograr es utilizando las respuestas de los estudiantes para involucrarlos en el discurso (Black y Wiliam 2009). Si el movimiento de *feedback* se usa no sólo para evaluar sino también para facilitar el aprendizaje de los estudiantes, entonces el significado puede ser co-construido y las oportunidades de aprendizaje pueden ser mayores (Wells 1993). Este estudio pretende ilustrar una gama de funciones diferentes para las que los

profesores pueden utilizar el tercer turno y explicar las diferentes implicaciones que cada uno de ellos tiene para la calidad de la interacción y las oportunidades de aprendizaje.

En el discurso del aula AICLE, la interacción es una oportunidad tanto para el desarrollo de la segunda lengua como para el aprendizaje del contenido. Por lo tanto, si los estudiantes están involucrados en la interacción, se puede desarrollar el conocimiento de lenguas extranjeras y contenido (Morton 2012). Al igual que en las aulas de inglés como lengua extranjera (EFL), parece que las secuencias IRF son muy frecuentes en aulas AICLE (Dalton-Puffer 2007). Sin embargo, en comparación con las clases de EFL, los IRF en AICLE parecen dar lugar a una enseñanza más dialógica, con respuestas más largas de los estudiantes y una variedad de funciones realizadas en el tercer turno (Nikula 2007). El tercer turno es crucial en cualquier tipo de aula, pero aún más en el caso de AICLE: los maestros necesitan una variedad de estrategias interaccionales para ayudar a los estudiantes a comprender y aprender conceptos co-construidos en un idioma extranjero (Evnitskaya 2012). En este estudio, se prueba que el tercer turno es útil para que los docentes evalúen el conocimiento de la lengua y del contenido de los estudiantes, y para guiar el aprendizaje de contenido y lengua por parte del alumnado.

En el discurso del aula, los largos intercambios de IRF dan lugar a episodios distintos. Por tanto, durante las lecciones surgen episodios diferentes, como muchos estudios ya han puesto de relieve y analizado (Frölich, Spada y Allen 1985; Bloome et al. 2009; Snell y Lefstein 2011; Berg 2009). Como se plantea en esta disertación, los diferentes tipos de episodios pueden estar más o menos alineados con una pedagogía AfL, dependiendo de lo que se está haciendo y cuál es el objetivo del episodio. Si el tipo de episodio está diferencialmente alineado con AfL, entonces también tiene un impacto diferencial en el tipo de preguntas y *feedback* utilizados por el profesor. Del mismo modo, el tipo de episodio también puede afectar la forma en la que los estudiantes están involucrados en el discurso.

1.4 El estudio: Objetivos, preguntas de investigación e hipótesis

Como se ha señalado anteriormente, la evaluación en AICLE es una de las áreas que más investigación necesita, especialmente cuando se trata de la relación entre la evaluación formativa y AICLE (Llinares, Morton y Whittaker 2012). Como gran parte del trabajo formativo se realiza a través de la interacción profesor-alumno (Leung y Mohan 2004), el objetivo principal que intento lograr en esta tesis es ofrecer una explicación detallada de cómo los maestros construyen la Evaluación para el Aprendizaje en las aulas AICLE y cómo esto afecta a la participación de los estudiantes. En otras palabras, averiguar cómo el discurso

moldea AfL a través de diversos patrones interaccionales y estrategias del profesor. Las preguntas de investigación relacionadas con este objetivo principal son las siguientes:

1. ¿Cuáles son las características y estrategias interaccionales que caracterizan el discurso del AfL en las aulas AICLE de primaria?
2. ¿Son estos patrones específicos para las clases AfL o también se encuentran en clases similares donde AfL no se implementa?
3. ¿Existen diferencias en estas características y estrategias de interacción que caracterizan el discurso del AfL en diferentes asignaturas? ¿Hay diferencias en el discurso del mismo profesor en diferentes materias?
4. ¿Hasta qué punto afectan las estrategias de AfL a la participación de los estudiantes en el discurso de clase?

Los objetivos de esta investigación son:

1. Caracterizar cómo se construye el discurso AfL en los diferentes niveles de Primaria y en diferentes colegios, a través de materias diferentes y con diferentes profesores.
2. Comparar las características discursivas de AfL en los colegios que implementan este enfoque y los patrones discursivos utilizados en las escuelas AICLE similares que no lo implementan específicamente.
3. Comparar los patrones de interacción de los profesores en diferentes asignaturas.
4. Descubrir cómo las estrategias AfL afectan la participación de los estudiantes en el discurso.

A la luz de las preguntas de investigación y los objetivos expresados anteriormente, las hipótesis correspondientes con las que trabajaré son las siguientes:

1. La primera hipótesis es que AfL en AICLE se construye a través de un tipo específico de discurso, con el uso de ciertos tipos de preguntas, ciertos episodios y ciertos tipos de *feedback* por parte del profesor.
2. La segunda hipótesis articula que habrá diferencias en los patrones discursivos ofrecidos por los profesores en las escuelas AfL y aquéllos ofrecidos por los profesores en los colegios No-AfL

   Las dos primeras hipótesis se basan en la teoría de AfL, que pone de relieve la importancia de una buena formulación de preguntas por parte de los profesores, el abandono de las preguntas fácticas y el uso de preguntas de alto nivel, y la importancia de la calidad del *feedback* del docente, alejándose de la mera evaluación e intentando mediar en los procesos de aprendizaje del alumnado (Black y Wiliam 1998a, b; Black et al. 2003; Harrison y Howard 2009).

3. La tercera hipótesis establece que a) habrá diferencias en los patrones de interacción utilizados en las diferentes asignaturas y b) se encontrarán diferencias en el discurso del mismo profesor enseñando materias diferentes. Esto se basa en la evidencia de que AfL varía dependiendo de diferentes factores, siendo uno de ellos que el mismo profesor puede invocar diferentes supuestos y principios pedagógicos en diferentes contextos (asignaturas) y con diferentes estudiantes (Black et al. 2004; Wiliam 2006; Black y Wiliam 1998a; Black y Wiliam 2009; Hodgen y Marshall 2005; Leung 2004, Torrance y Pryor 1998).
4. La cuarta hipótesis afirma que las estrategias AfL influirán positivamente en la participación y las contribuciones de los estudiantes. Como consecuencia, se plantea la hipótesis de que habrá diferencias en los tipos de respuestas dadas por los estudiantes en las escuelas AfL y No-AfL.

1.5 El estudio: Datos y metodología

La evaluación formativa de los profesores ha sido apoyada por los agentes políticos en diversos países como Inglaterra, Australia y Hong Kong. Ése no es el caso en España, donde sólo los colegios que forman parte del proyecto British / MEC implementan conscientemente este tipo de evaluación. La mitad de las escuelas en las que se realizó el presente estudio pertenecen a este proyecto, y la otra mitad pertenecen al proyecto de la Comunidad de Madrid, en el que no se está aplicando la Evaluación para el Aprendizaje. El hecho de que los datos provengan de colegios AICLE en el nivel de primaria es una contribución importante de esta tesis, ya que la mayoría de las investigaciones sobre interacción (y la investigación AICLE en general) se han centrado en la educación secundaria (Dalton-Puffer 2007; Llinares et al. 2012; Llinares y Whittaker 2009; Nikula 2005, 2008; Moore 2011), pero pocos estudios se centran en el nivel universitario (Dafouz Milne y Núñez Perucha 2010; Smit 2010) o en los niveles primarios (Llinares y Lyster 2014; Pastrana 2010; Serra 2007).

Los datos del presente estudio incluyen sesiones de aula de cuatro colegios. En cada colegio se grabaron sesiones en dos asignaturas impartidas por el mismo profesor. Los niveles grabados son 5º y 6º de Primaria (los estudiantes tienen 10-11 y 11-12 años, respectivamente) y las asignaturas grabadas son Ciencias, Artes, Drama y Ciudadanía. Ha habido dos rondas de grabaciones en cada escuela en las dos materias diferentes, una al principio del año escolar y la otra al final. De esta forma, los datos eran más fiables, ya que se podían controlar las variables tema de la unidad y el tiempo de grabación, evitando distorsiones de resultados porque a los estudiantes les gustara o no les gustara un tema específico, o porque los estudiantes estaban más

o menos familiarizados con su maestro, su estilo de enseñanza y/o sus compañeros de clase. Del mismo modo, se podrían detectar diferencias (si las hay) entre la co-construcción del discurso en estos dos momentos diferentes a lo largo del año escolar. Se grabaron unidades didácticas completas en cada asignatura, lo que significa que hay dos unidades didácticas completas para cada una de las disciplinas registradas en cada colegio. Las unidades didácticas variaron de dos a cinco sesiones. En total, se han utilizado 44 sesiones de aula para esta tesis, que asciende a un total de aproximadamente 50 horas y un número total de alrededor de 350.000 - 400.000 palabras.

Se analizaron todas las interacciones profesor-alumno ya que, siempre que hay interacción, existe la posibilidad de que el AfL se implemente. Cada sesión de aula se dividió en episodios, "que consiste en toda la charla producida en llevar a cabo una sola actividad o una de sus tareas constitutivas" (Nassaji y Wells 2000: 383). Los episodios se clasificaron según el propósito de la actividad. En otras palabras, de acuerdo con lo que el profesor y los estudiantes estaban haciendo en ellos.

En este estudio, la unidad principal de análisis es el *Exchange*. Sus diferentes componentes (Iniciación-Respuesta-*Feedback*) se analizaron como sigue:

1. Iniciaciones: se clasificaron de acuerdo a quién inicia el intercambio: profesor o alumno. Cuando el profesor inicia y esa iniciación es una pregunta, ésta fue clasificada siguiendo la tipología de Dalton-Puffer (2007) para las aulas AICLE: preguntas sobre hechos y datos, preguntas sobre explicaciones, preguntas sobre razones, preguntas de opinión y preguntas meta-cognitivas (ver la clasificación en el Capítulo 4). También se analizaron las iniciaciones de los estudiantes, prestando especial atención al tipo de contribución que hacían: hacer una pregunta, expresar una opinión personal, exponer un hecho, explicar, argumentar, etc. Esto podría proporcionar información sobre si el tipo de enseñanza es más dialógico o autoritario y cómo esto se correlaciona o no con la esencia de la evaluación formativa.
2. Respuestas: como forma de medir la longitud y la complejidad de las respuestas de los estudiantes, éstas se dividieron en respuestas mínimas, respuestas truncadas y respuestas formadas por unidad-T (a su vez divididas en un sintagma, una proposición y más de una proposición). Las respuestas se analizaron y compararon teniendo en cuenta el tipo de clase (AfL o no), el tipo de pregunta previamente planteada por el profesor y la asignatura.
3. *Feedback*: los turnos de *feedback* de los docentes son otro elemento clave de la teoría AfL (ver por ejemplo Black y Wiliam 1998a; Torrance y Pryor 2001; Wiliam et al.

2004; Sadler 1998). El objetivo de esta investigación fue crear una taxonomía que describiera los diferentes y diversos tipos de *feedback* que utilizan los maestros AICLE, con especial atención a aquellos tipos que claramente ayudan a construir AfL.

1.6 Visión general de la tesis

La presente disertación se divide en nueve capítulos. Este primer capítulo es la introducción, que trata de establecer el propósito y el alcance del estudio, así como los objetivos, las preguntas de investigación y las hipótesis.

El segundo capítulo está enteramente dedicado a AICLE como modelo educativo. Se exploran varios aspectos de este modelo, como qué es, sus metodologías precursoras, su situación en España y más concretamente en Madrid, la metodología AICLE, las líneas de investigación y los resultados del aprendizaje.

El tercer capítulo trata del marco teórico necesario para comprender, interpretar y llevar a cabo esta investigación. Como se señaló anteriormente, en esta sección hay dos ejes principales: el discurso en el aula y la Evaluación para el Aprendizaje. Dentro del discurso en el aula, los patrones IRF son el enfoque clave, así como el discurso y la interacción en aulas AICLE. Además, se explorarán los efectos de los diferentes tipos de *feedback* sobre el aprendizaje de idiomas por parte de los estudiantes. En la sección de AfL, sus características se explican de una manera detallada, especialmente concentrándose en el *feedback* y la autoevaluación y co-evaluación. Por último, se explora el papel del discurso en el aula en la Evaluación para el Aprendizaje.

El cuarto capítulo presenta todos los datos que se han utilizado para este estudio, así como la forma en que los datos han sido tratados y analizados.

Los capítulos 5, 6, 7 y 8 se centran en los resultados. En todos ellos se incluyen ejemplos del corpus para ilustrar los resultados obtenidos del análisis. El capítulo 5 se centra en los resultados obtenidos del análisis de episodios. El capítulo 6 se centra en el análisis de los tipos de preguntas de los profesores, comenzando así el análisis de las secuencias IRF. Para continuar con el análisis de los IRF, las respuestas de los estudiantes se describen y discuten en el capítulo 7. En último lugar, el capítulo 8 se centra en el *feedback* ofrecido por parte de los profesores.

Aunque cada capítulo de resultados incluye una sección de discusión, el capítulo final está dedicado a discutir los resultados a un nivel más global: los principales resultados del estudio se ponen de relieve a la luz de los objetivos iniciales establecidos para el estudio. Además, se discute el papel de la interacción en las aulas AICLE AfL. Este capítulo también proporciona

una propuesta para un modelo de prácticas de discurso de AfL en las aulas AICLE que permitirá a los profesores evaluar las brechas de aprendizaje del contenido y del idioma extranjero por parte del alumnado y ajustar su enseñanza para cerrar dichas brechas y así mejorar el aprendizaje de los estudiantes. Las aplicaciones pedagógicas, las limitaciones del estudio y las investigaciones adicionales también se tratan en este capítulo. Por último, algunas observaciones finales complementan todo el trabajo presentado a lo largo de esta tesis.

1.7 Resumen del capítulo

El presente capítulo es una guía de lo que se puede encontrar en esta disertación. En primer lugar, se ha ofrecido una visión general del propósito y el alcance del estudio, presentando los diferentes marcos y por qué esta tesis puede ser interesante y contribuir con nuevos hallazgos. En segundo lugar, se han presentado y conectado brevemente las diferentes perspectivas teóricas utilizadas para este estudio y se han planteado algunas cuestiones importantes (véanse los capítulos 2 y 3 para el marco teórico ampliado). En tercer lugar, se han formulado los objetivos, las preguntas de investigación y las hipótesis de la tesis. A continuación, los datos y la metodología utilizados se han explicado de forma concisa (véase el capítulo 4 para una descripción detallada). Finalmente, este capítulo ha presentado la organización de la tesis, proporcionando una breve descripción de lo que se encuentra en cada una de sus diferentes secciones.

2 DISCUSIÓN GENERAL Y CONCLUSIONES

Los capítulos 5-8 presentaron los resultados de cada uno de los patrones de interacción que se han investigado en el presente estudio (episodios, preguntas, respuestas y *feedback*), incluyendo una discusión para cada uno de estos patrones. Estos capítulos también proporcionaron un análisis comparativo dentro de cada uno de estos patrones entre las aulas de AfL / No-AfL, asignaturas y profesores individuales. En este capítulo final, presentaré una discusión general que girará en torno a las preguntas e hipótesis de investigación planteadas en el capítulo introductorio. A la luz de estos resultados, el capítulo pasará a discutir el papel de la interacción en las aulas AICLE y AfL, lo que conducirá a una propuesta final de un modelo interaccional que podría ser utilizado para implementar el AfL en este tipo de clases. El presente capítulo también destacará las implicaciones pedagógicas que pueden aplicarse a las pedagogías de AICLE o de AfL. El capítulo concluirá con un reconocimiento de las limitaciones del estudio y

la formulación de ideas para futuros estudios que podrían ser interesantes y que podrían completar la presente investigación.

2.1 Objetivos y conclusiones iniciales

Esta sección presentará los principales hallazgos en relación con las preguntas iniciales de investigación y las hipótesis de investigación. El objetivo principal era descubrir cómo el discurso da forma al AfL en las aulas AICLE a través de diversos patrones de interacción. Las preguntas específicas de investigación fueron las siguientes:

1. ¿Cuáles son las características y estrategias interaccionales que caracterizan el discurso del AfL en las aulas AICLE de primaria?
2. ¿Son estos patrones específicos para las clases AfL o también se encuentran en clases similares donde AfL no se implementa?
3. ¿Existen diferencias en estas características y estrategias de interacción que caracterizan el discurso de AfL en diferentes asignaturas? ¿Hay diferencias en el discurso del mismo profesor en diferentes materias?
4. ¿Hasta qué punto afectan las estrategias de AfL a la participación de los estudiantes en el discurso de clase?

Las principales hipótesis relacionadas con estas preguntas fueron las siguientes:

1. El AfL se construye a través del discurso de clase por el profesor y los estudiantes a través del uso de ciertos tipos de episodios, de preguntas y de *feedback*.
2. Habrá diferencias en los tipos de episodios, en los tipos de preguntas y en los tipos de *feedback* utilizados por los profesores en los colegios AfL y aquéllos usados por los profesores en los colegios No-AfL.
3. Las diferencias se encontrarán a) en los patrones de interacción utilizados en las diferentes asignaturas, y b) en el discurso del mismo profesor enseñando materias diferentes.
4. Las estrategias AfL influirán positivamente en la participación y contribuciones de los estudiantes. Como consecuencia, encontraremos también diferencias en los tipos de respuestas dadas por los estudiantes en los colegios AfL y los No-AfL.

En primer lugar, en relación a la primera pregunta de investigación, se ha analizado la interacción en las clases AICLE de primaria impartidas por diferentes profesores y en diferentes materias con el propósito de identificar características que hagan esta interacción formativa. La primera de las características analizadas fueron las preguntas de los profesores. Tal y como se

planteó (ver hipótesis 1), el estudio ha demostrado que hay ciertos tipos de preguntas que se alinean con una pedagogía AfL y otras que no lo hacen. Las preguntas de opiniones, de razones, las preguntas meta-cognitivas y las meta-preguntas se alinean con la pedagogía AfL porque buscan explorar los procesos de pensamiento y aprendizaje de los estudiantes, como se muestra en las respuestas que éstos ofrecen. Por el contrario, las preguntas sobre hechos o datos no se alinean con tal pedagogía, ya que se centran en recordar los hechos más que en extender el razonamiento y la comprensión de los estudiantes (Black et al. 2003; Alexander 2004; Heritage 2007; Ruiz-Primo y Furtak 2007; Wragg y Brown 2001). El segundo rasgo analizado fue el *feedback* de los profesores. Los resultados mostraron que, en las lecciones de AfL, la expansión, re-orientación o meta-*feedback* eran frecuentes, en convergencia con una metodología AfL, que anima a los estudiantes a reflexionar sobre sus procesos de aprendizaje o pensamiento. Estos tipos de *feedback* toman en cuenta las respuestas de los estudiantes y facilitan la implicación de los mismos en la interacción, por lo que están más alineados con AfL. A su vez, la evaluación a menudo cierra el intercambio para dar paso a uno nuevo, evitando así que los estudiantes vuelvan a coger el turno después (Zhang Waring 2008, Wragg y Brown 2001, Black y Wiliam 1998a, b, Alexander 2004). Por otro lado, los tipos de *feedback* como la expansión, la re-orientación o el meta-*feedback* promueven que los alumnos puedan continuar con el intercambio y / o los empujan a reflexionar sobre sus procesos de aprendizaje y pensamiento (Ruiz-Primo y Furtak, 2006; Wiliam 1998a, b, Alexander, 2004). Del mismo modo, ciertos tipos de episodios son más característicos de un enfoque AfL: *exponer los objetivos de la lección*, *explicar las notas* y la *autoevaluación y co-evaluación*. Estos tipos de episodios parecen ofrecer un buen espacio para que los profesores hagan meta-preguntas a los estudiantes y les ofrezcan meta-*feedback*.

En segundo lugar, en respuesta a la segunda pregunta e hipótesis de investigación, hubo diferencias significativas entre las clases en las que se implementó AfL y en las que no se implementó: en las primeras, los docentes usaron significativamente más de los tipos de preguntas y *feedback* que están más asociados con una pedagogía de la evaluación formativa, es decir, más preguntas de opiniones, de razones, meta-cognitivas y meta-preguntas, y más expansión y meta-*feedback*.

En lo que respecta a los episodios, los que se alinean con una pedagogía AfL (mencionados anteriormente) se encuentran casi exclusivamente en las escuelas AfL. Sin embargo, también se observaron algunos puntos comunes entre los tipos de colegios: la *discusión de toda la clase* y el *manejo y rutinas de clase* fueron los episodios más frecuentes en ambos tipos. La *revisión* y *explicación de la actividad / tarea* también están presentes con porcentajes similares en ambos tipos de escuelas. Podríamos argumentar, por tanto, que los episodios característicos de la

pedagogía del aula en general eran más frecuentes, sin importar si la escuela estaba implementando AfL o no.

En cuanto a la tercera pregunta e hipótesis de investigación, también se encontraron diferencias significativas en los patrones del discurso en las diferentes asignaturas. En el caso de Ciencias, tanto las preguntas sobre hechos o datos como las preguntas sobre explicaciones predominan, así como la evaluación y la expansión, lo que indica que el contenido se trata y se pregunta frecuentemente como hechos, y los estudiantes a veces son alentados a elaborar y explicar ideas. En Ciudadanía, las preguntas y el *feedback* utilizadas por los profesores no sólo versan sobre hechos, sino que también buscan animar a los estudiantes a discutir puntos de vista, elaborar ideas y reflexionar sobre los procesos de aprendizaje. En las lecciones de Arte, los resultados sobre los tipos de preguntas y los tipos de *feedback* son de alguna manera contradictorios: mientras que la mayoría de las preguntas son sobre hechos y datos, la evaluación en el turno de *feedback* no es tan frecuente y el *meta-feedback*, la expansión y la re-orientación son prominentes. En cuanto a las clases de Drama, los puntos de vista de los estudiantes, las opiniones personales y las reflexiones sobre la evaluación y el aprendizaje se fomentan a través de preguntas de opinión, preguntas meta-cognitivas y meta-preguntas. Del mismo modo, aparte de la evaluación, el *feedback* en las lecciones de Drama también incluye la expansión, la re-orientación y el meta-*feedback*, lo cual sigue la misma línea que los tipos de preguntas encontradas. Con estos resultados en mente, parece que las Ciencias, al menos en el contexto de las clases AICLE de primaria analizadas, es la asignatura menos favorable para las técnicas de AfL. Estos resultados, por supuesto, tendrían que contrastarse entre niveles educativos y áreas geográficas. En un estudio contrastivo reciente sobre la evaluación y el lenguaje evaluativo por parte de los estudiantes finlandeses y españoles en las aulas AICLE de biología en la escuela secundaria, Llinares y Nikula (2016) observaron diferencias en los papeles de participación de los estudiantes españoles y finlandeses asignados por los profesores y su efecto sobre su uso de la lengua, el acercamiento al contenido, y la participación general en el aula.

También como parte de la tercera pregunta e hipótesis de investigación, el análisis reveló que el hecho de que todos los profesores mostraran diferencias en cuanto a tipos de preguntas y *feedback* en sus dos asignaturas respalda el efecto de la asignatura en los patrones de interacción utilizados. Sin embargo, como no todos los profesores enseñaban las mismas dos asignaturas, no podemos afirmar categóricamente que las diferencias se deban exclusivamente a las especificidades de la materia. La variable profesor podría estar jugando muy bien su papel también. De hecho, parece ser que ambas variables (maestro y asignatura) desempeñan un papel importante. El único caso claro en el que los tipos de *feedback* parecen afectados principalmente por la materia son las Artes, ya que las actividades prácticas llevaron a los profesores a dar

comentarios sobre la forma en que los estudiantes estaban haciendo el trabajo (meta-*feedback*), a pesar de que los dos profesores que enseñaban Artes eran No-AfL y sus porcentajes totales de meta-*feedback* fueron significativamente menores que el de los profesores AfL. Se esperaba que los profesores AfL se alinearan con un discurso de AfL sin importar qué asignatura estaban enseñando. Al contrario de lo que se esperaba, tanto los docentes AfL como los No-AfL muestran diferencias en su uso de preguntas y *feedback* en diferentes materias. En consecuencia, parece que el tipo de asignatura parece desempeñar un papel más importante que la variable AfL o No-AfL. Los episodios también fueron analizados en los diferentes profesores y materias. Como ya se señaló en el caso de los tipos de preguntas y de *feedback*, en el caso de los episodios, parece que tanto la variable del estilo de enseñanza como la variable asignatura juegan un papel en el tipo de episodios utilizados. Esto es más evidente en el caso de los profesores de AfL, ya que los episodios relacionados con una pedagogía AfL están presentes en todas sus materias.

En relación con la cuarta pregunta de investigación e hipótesis de este estudio, los resultados muestran que los tipos de preguntas y *feedback* que se alinean con la metodología AfL tienen un impacto positivo en la participación y las contribuciones de los estudiantes. En primer lugar, las respuestas de los estudiantes a las preguntas que se alinean con una pedagogía AfL eran más complejas, conteniendo al menos una unidad T que, en muchos casos, estaba compuesta por más de una proposición, implicando coordinación o subordinación. Del mismo modo, los resultados de esta investigación indican que era más probable que los estudiantes tuvieran oportunidades de recuperar el turno de palabra después del *feedback* de los profesores cuando este *feedback* era formativo en lugar de evaluativo.

2.2 El papel de la interacción en las aulas AICLE y AFL

La calidad de la interacción está en el centro de una buena pedagogía (Black y Wiliam 1998a). Esta investigación ha mostrado cómo diferentes tipos de patrones de interacción pueden conducir a diferentes niveles de participación de los estudiantes, lo que podría tener diferentes impactos en el aprendizaje de los estudiantes. Para proporcionar más apoyo a este respecto, en este estudio hemos demostrado que algunos tipos de preguntas y *feedback* en los patrones IRF pueden conducir a un tipo de interacción mucho más dialógico, dando lugar a unos turnos de los alumnos más largos y más complejos. El presente estudio ha mostrado cómo ciertos tipos de preguntas y *feedback* en los IRF fomentan el pensamiento de los estudiantes (a través del uso de preguntas meta-cognitivas, meta-preguntas, preguntas sobre razones, expansión y re-orientación), fomentan que los estudiantes persigan sus propias ideas (permitiéndoles iniciar el

discurso y hacer sus propias preguntas), así como también fomenta que los estudiantes reflexionen sobre el aprendizaje y la evaluación (a través de meta-preguntas y meta-*feedback*). Este tipo de interacción fue más frecuente en las aulas AICLE en las que se implementó AfL. Sin embargo, como se esperaba en cualquier contexto del aula, IREs (Iniciación-Respuesta-Evaluación) también aparecieron en las aulas AfL. Los IRE están vinculados a enfoques autoritarios donde se presta atención a una sola voz y punto de vista, sin explorar otras ideas (Mortimer y Scott 2003), y asociados con la visión de la enseñanza como un proceso de transmisión (Barnes 'Transmission Model', 1975), en el que el profesor es un mero transmisor de contenido que transmite información a los estudiantes. En las aulas No-AfL, la interacción fue más cercana a la recitación, que sigue siendo el modo más común de enseñar, a pesar de los beneficios que la enseñanza receptiva / contingente / dialógica tiene para el aprendizaje (Alexander 2004; Mortimer y Scott 2003). En la recitación predominan las preguntas sobre hechos y datos y las preguntas de razones, meta-cognitivas y meta-preguntas son escasas. Lejos de descartar IREs y patrones de recitación, la propuesta en este estudio es que los tipos de interacción de AfL sean integrados en cualquier clase, pero particularmente en aulas AICLE, donde quizá se necesiten más específicamente las oportunidades de uso de la lengua (para la práctica de la lengua meta) y diferentes maneras de abordar el contenido (para una mejor comprensión del contenido enseñado en una lengua extranjera).

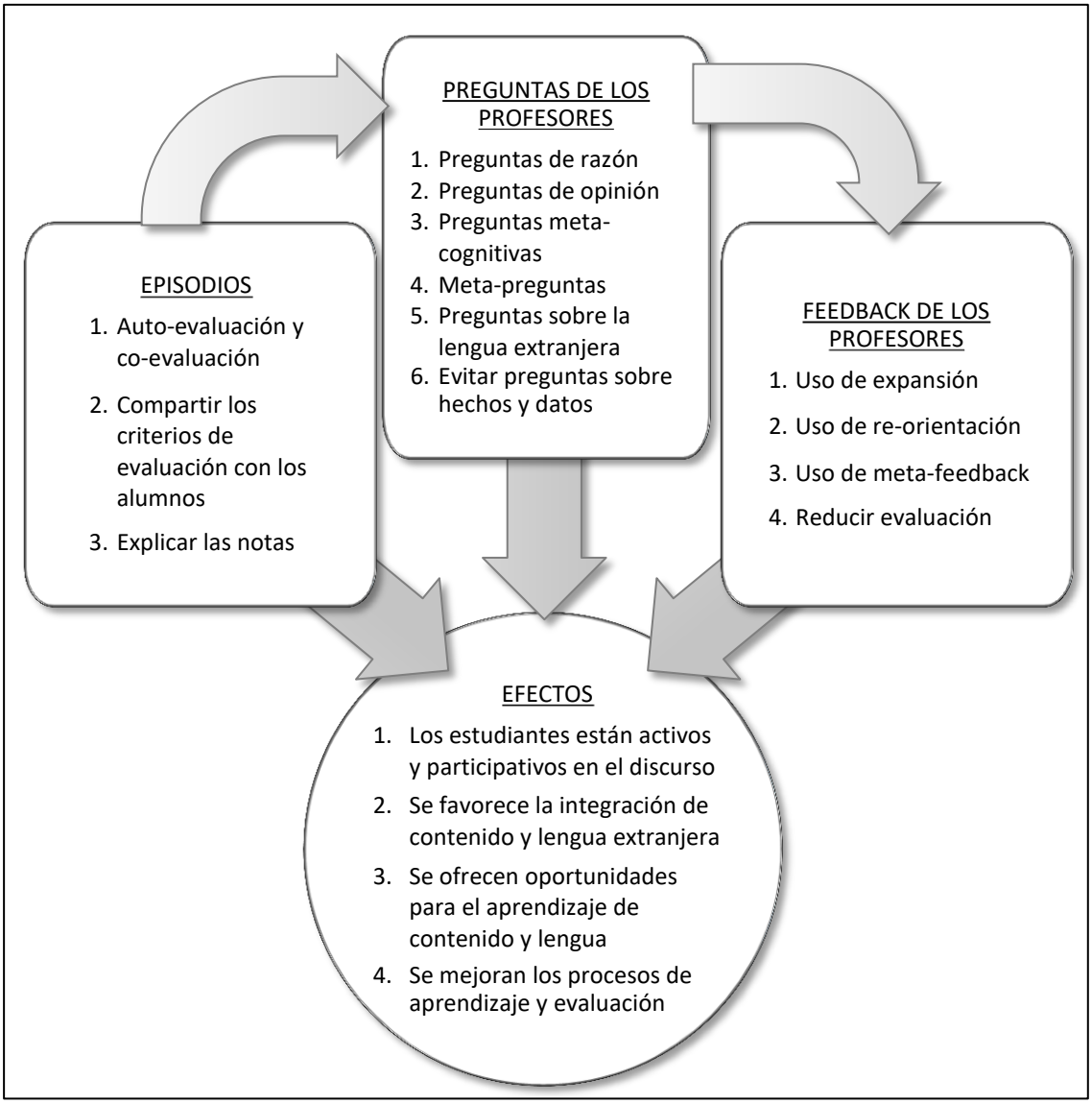
El presente estudio ha demostrado que las discusiones de toda la clase son el episodio más frecuente en los datos utilizados para esta investigación (junto con los episodios de manejo de clase), tanto en las aulas AfL como en las No-AfL. Este formato de interacción es la regla en la mayoría de las clases (Lyster 2007: 87, Lyster y Mori 2006, Dalton-Puffer 2006, Fazio y Lyster 1998, Hiebert 1999, Alexander 2004), incluyendo aulas AICLE (Nikula et al. 2013). Estos episodios son muy importantes para desarrollar un enfoque de investigación en el aula (Nassaji y Wells 2000). Sin embargo, el estudio también ha revelado que la utilización de algunos tipos de episodios, como la *explicación de las notas* y la *autoevaluación y co-evaluación*, también pueden ser útiles no sólo cuando los profesores de AICLE quieren implementar específicamente AfL, sino que en general ofrecen un buen contexto para que se produzcan meta-preguntas y meta-*feedback*. De esta manera, el aprendizaje de los estudiantes y su conocimiento de ese aprendizaje pueden ofrecer información crucial para los docentes, lo que les permite ajustar su enseñanza y decidir sobre los próximos pasos de enseñanza más adecuados (Harrison y Howard 2009; Black y Wiliam 1998b). Además, proporcionan excelentes oportunidades para el "lenguaje a través del aprendizaje" (Coyle 2010), el uso del lenguaje evaluativo y un enfoque evaluativo del contenido, que son clave para que los alumnos AICLE integren el aprendizaje de contenido y lengua (Morton y Llinares 2016; McCabe y Whittaker 2017).

Con respecto al tipo de interacción que se encuentra en diferentes materias, el presente estudio también ha revelado que el tipo de asignatura juega un papel cuando se trata de la producción lingüística de los estudiantes. Específicamente en las clases de lengua extranjera, la producción de los estudiantes fue más compleja en Ciudadanía, ya que es una materia que ofrece espacio para que los estudiantes exploren sus propias opiniones y reflexiones. Esto se contrapone con las Ciencias, donde es más frecuente recordar los hechos y dar explicaciones a los fenómenos científicos.

Finalmente, aunque el aprendizaje no ha sido tratado abiertamente en este estudio, los resultados han demostrado que la participación en la interacción proporciona oportunidades para la comprensión y el aprendizaje. (Dalton-Puffer y Nikula 2006; Dalton-Puffer y Smit 2007, Dalton-Puffer 2009). Se ha demostrado que las aulas AfL representan un contexto más favorable para promover el aprendizaje, ya que los estudiantes participan más y con turnos más largos, a menudo inician el discurso, se les pide que expliquen y argumenten sus ideas y puntos de vista y se les anima a expandir sus contribuciones. Como sostiene Barnes (1975), diferentes tipos de comunicación conducen a diferentes tipos de aprendizaje. De esta manera, las aulas en las que se hacen muchas preguntas sobre hechos y datos desencadenan la memorización y el aprendizaje por repetición y memoria, en oposición al razonamiento y el aprendizaje profundo, alentados cuando los profesores hacen otros tipo de preguntas (preguntas sobre razones, explicaciones, preguntas meta-cognitivas…). Esto se relaciona con el concepto de competencia interaccional en el aula (CIC) de Walsh (2006; Seedhouse y Walsh 2010) y la identificación de decisiones instantáneas hechas por el profesor y los estudiantes en interacción para crear espacios de aprendizaje (ajustar la interacción para satisfacer las necesidades y objetivos). Como se muestra en este estudio, la creación de estos espacios de aprendizaje puede hacerse a través de preguntas de los profesores que apuntan al razonamiento y al aprendizaje más profundo, permitiendo que los estudiantes participen y contribuyan al discurso y ofreciendo expansión, re-orientación y meta-*feedback* en el tercer movimiento.

2.3 Propuesta para un modelo de prácticas discursivas AfL en las aulas AICLE

El siguiente modelo es una propuesta nacida de la presente investigación basada en los diferentes análisis realizados sobre la interacción en el aula en contextos AfL y No-AfL. Esta guía puede ser útil no sólo para la implementación de prácticas exitosas AfL en contextos AICLE, sino también para incrementar las oportunidades de prácticas más exitosas de interacción de lengua y contenido en las aulas AICLE en general.



Modelo para implementar el AfL interactivo en las aulas AICLE.

Una implementación exitosa de AfL en AICLE requiere que los maestros hagan un uso especial de ciertos tipos de preguntas: preguntas sobre razones, preguntas meta-cognitivas y meta-preguntas. Como se ve en este estudio, con estos tipos de preguntas, los profesores pueden hacer que los estudiantes razonen y piensen. A su vez, las preguntas sobre hechos y datos, que son necesarias en la escuela, tienen que ser reducidas, ya que la participación de los estudiantes cuando se les pregunta ha demostrado ser menor. Además de los tipos de preguntas antes mencionados, los profesores también pueden decidir utilizar las preguntas sobre la lengua cuando, a la luz de las contribuciones de los estudiantes, sientan que algunos aspectos lingüísticos merecen especial atención. En lo que respecta a los tipos de *feedback*, los profesores deben tratar de utilizar expansión, meta-*feedback* y re-orientación siempre que las

intervenciones de los estudiantes desvelen conceptos erróneos. Los maestros pueden redirigir el pensamiento de los estudiantes a través del uso de *recasts* o *prompts*. Los resultados actuales han demostrado que, a través del uso de estos tipos de *feedback*, los profesores animan a los estudiantes a reflexionar sobre sus procesos de aprendizaje y evaluación, así como también amplían y expanden las contribuciones de los mismos, lo cual facilita el aprendizaje, tanto de contenido como de la lengua extranjera. Como se demostró en esta investigación, a través del uso de los tipos de pregunta y *feedback* antes mencionados, los estudiantes están activos y participan en la interacción y el proceso de aprendizaje. De la misma manera, si los profesores utilizan ese tipo de preguntas y *feedback*, las contribuciones de los estudiantes, como se muestra en este estudio, son largas y complejas. Esta participación activa en el discurso también ofrece al alumnado la oportunidad de usar el idioma extranjero con un propósito. Como resultado de la implementación del modelo propuesto, esta investigación ha demostrado que los estudiantes no sólo pueden participar en el discurso con turnos largos sino también iniciar el discurso y realizar una serie de funciones diferentes al hacerlo. Esto no sólo hace que la interacción sea más simétrica, sino que también permite a los estudiantes encontrar contextos significativos y relevantes que pueden ayudarles en su aprendizaje de contenido y lengua extranjera. Para este propósito, los episodios como la *explicación de las notas*, la *autoevaluación y la evaluación de los compañeros* y *compartir los criterios de evaluación con los estudiantes* son especialmente relevantes para que éstos participen activamente en sus propios procesos de aprendizaje y evaluación. Por lo tanto, proporcionan un contexto apropiado para que los profesores hagan meta-preguntas y proporcionen met*a-feedback* y que los alumnos se apropien de su proceso de aprendizaje.

2.4 Aplicaciones pedagógicas

Se pueden extraer diferentes implicaciones pedagógicas del presente estudio. Una de las principales implicaciones es que la naturaleza receptiva de AfL radica en la calidad de la interacción, los tipos de preguntas y el *feedback* de los profesores, y la participación de los estudiantes en el discurso. Estas características son particularmente relevantes en las aulas AICLE para la integración de contenido y lengua, y para la doble mediación que necesita existir en este tipo de clases (Gibbons 2003). Además, en AICLE es muy importante elevar los procesos meta-cognitivos de los estudiantes, es decir, hacerlos conscientes de cómo aprenden a aprender (Coyle 2006). Esto está estrechamente relacionado con AfL, ya que una característica clave de este enfoque es que los estudiantes participan activamente en sus procesos de aprendizaje y son conscientes de ellos (Black y Wiliam 1998a, b, Grupo de Reforma de

Evaluación 2002). Es por eso que las meta-preguntas y el meta-*feedback* serían muy importantes en cualquier aula AICLE, independientemente de que se esté implementando o no AfL. Sin embargo, esto no sucede con tanta frecuencia cuando AfL no está presente, como lo demuestra esta investigación.

Como se muestra en este estudio, la implementación de AfL aumenta significativamente la calidad de la interacción en el aula, un componente clave de cualquier tipo de aprendizaje, más aún en el caso de las aulas AICLE, donde se espera que el contenido y la lengua sean aprendidos de manera integrada. Por tanto, es necesario implementar programas de formación de profesores en los que el enfoque interactivo de AfL explorado en esta tesis sea dado a conocer a los profesores AICLE. Como muchos autores han afirmado, en los programas de AfL los profesores necesitan renegociar el contrato de aprendizaje (Perrenoud 1991; Black et al. 2003; Harrison y Howard 2009; Heritage 2010), lo cual significa que la responsabilidad se transmite a los estudiantes y que el objetivo de los docentes no es sólo transmitir conocimiento.

Los programas de formación de profesores de AfL en países como el Reino Unido han demostrado ser exitosos pero lentos (Black et al. 2003; Black et al. 2004; Black y Wiliam 2003; Wiliam et al. 2004). Para que estos programas tengan éxito, es primordial que la comunidad escolar trabaje en conjunto, que los maestros y los estudiantes tengan apoyo administrativo y que todos los agentes crean que la inversión y el esfuerzo valen la pena y que marcarán una diferencia en el aprendizaje futuro de los estudiantes. Muchas veces, la barrera principal son las tensiones y los choques existentes entre programas de enseñanza tradicionales e innovadores, así como el hecho de que la evaluación formativa se pone a menudo en competencia directa con la evaluación sumativa (Black y Wiliam, 1998a). En aras de la educación, no obstante, es necesario encontrar conciliaciones, y los dos tipos de evaluación deben ser complementarios (Llinares et al. 2012; Black et al. 2011; Black 2012).

Los programas de formación de profesores sobre la aplicación de AfL pueden ser beneficiosos para cualquier tipo de aula (como ya se ha demostrado en otros países, como el Reino Unido - véase Black y Wiliam 2003, Harrison y Howard 2009), pero tendrían un valor añadido en el caso de las aulas AICLE. La interacción es crucial para el aprendizaje de idiomas. Por lo tanto, cuanto mayor sea la calidad de la interacción, las oportunidades para que los estudiantes desarrollen su competencia en lenguas extranjeras serán mejores. Al mismo tiempo, la interacción en las aulas AICLE es también el medio a través del cual se desarrolla el aprendizaje del contenido. Como se muestra en este estudio, esta interacción puede ser más atractiva y significativa si se implementa AfL. En otras palabras, un enfoque AfL puede permitir que los estudiantes no sólo aprendan el contenido de una manera diferente, sino que también usen la

lengua extranjera en un rango más amplio de prácticas interaccionales centradas en el estudiante. Por último, dado que AfL se caracteriza por su capacidad de respuesta y por su contingencia, ayudaría a los profesores AICLE a proporcionar a los estudiantes el andamiaje adecuado a los dos niveles necesarios, contenido y lengua extranjera, así como a integrar ambos aspectos de manera significativa. Como dicen Llinares et al. (2012), tres características de AfL son esenciales en AICLE: AfL está planificada (por lo tanto, no sólo el conocimiento y los conceptos, sino también el lenguaje que los acompaña, deben planificarse e incluirse en la enseñanza, así como los géneros y los registros junto con su gramática y vocabulario); AfL es reactiva (se ajusta al proceso de enseñanza); AfL es recíproca (fomenta la autonomía del alumno, y eso puede conducir al desarrollo de la lengua extranjera).

Los programas de formación de profesores en AICLE no sólo tendrían que incluir un enfoque de la interacción en el aula en general, sino también una comprensión de cómo funciona la interacción en el aula en función de la asignatura. Como se ha demostrado en éste y otros estudios, la asignatura es una variable importante que debe tenerse en cuenta (Black et al., 2004; Wiliam 2006; Black y Wiliam 1998a; Black y Wiliam 2009; Hodgen y Marshall 2005). Esto se vuelve especialmente relevante en AICLE, ya que los alumnos tienen que aprender en un idioma extranjero los diferentes géneros asociados a las diferentes materias (Dale y Tanner 2012, Llinares et al. 2012; Llinares y Whittaker 2009). Como señalan Llinares y Pascual (2014), el discurso de clase puede ayudar a construir diferentes géneros, desde las complejas explicaciones científicas en Ciencias hasta el desarrollo de opiniones personales en Ciudadanía. Por lo tanto, la interacción en las aulas AICLE es crucial para que los estudiantes se familiaricen y aprendan los diferentes géneros asociados con diferentes materias escolares.

2.5 Limitaciones del estudio

Cada estudio tiene sus limitaciones, ya que su alcance debe ser acotado. Una de las limitaciones que se podría argumentar sobre esta investigación es que el aprendizaje no ha sido tratado como tal. Como señalan Mortimer y Scott (2003: 101), “los análisis se llevan a cabo, y los resultados se presentan, únicamente en términos de patrones de interacción, y el contenido real de lo que se está enseñando y aprendiendo no se considera como una característica significativa “. En este estudio se ha analizado la participación de los estudiantes y las respuestas a diferentes técnicas de AfL, y se han formulado conclusiones sobre las oportunidades de estas contribuciones para el aprendizaje de contenidos y de lengua extranjera. Aunque se espera que la participación de los estudiantes mejore el aprendizaje, el estudio no ha demostrado que las técnicas de AfL hayan mejorado realmente el aprendizaje de contenido y lengua por parte de los estudiantes. Como se

señaló anteriormente en este estudio, el aprendizaje es difícil de medir porque, si bien tiene lugar en la interacción social, también implica procesos cognitivos a los que no hay acceso (Seedhouse 2010; Ellis 2010). Por lo tanto, con la interacción en el aula sólo se pueden hacer inferencias de aprendizaje (Ellis 2010). Sin embargo, cada vez más estudios ponen de relieve el papel del aprendizaje en los contextos sociales y algunos de los ejemplos mostrados en este estudio han ilustrado momentos de aprendizaje. Es cierto que, en este estudio, el aprendizaje podría haber sido medido a través de pre y post-pruebas (por ejemplo, Ruiz-Primo y Furtak 2006). Sin embargo, las dimensiones de este estudio, incluyendo las unidades didácticas completas de diferentes asignaturas enseñadas por los mismos profesores en los colegios AfL y No-AfL, hubieran dificultado la creación de un pre y post-test para cada uno de los grupos / contextos.

Otra limitación del estudio es que no disponemos de datos de todos los docentes enseñando las mismas asignaturas y las mismas unidades didácticas, lo que habría hecho más fiable la comparabilidad entre ellos. Sin embargo, el acceso a los colegios es muy difícil, aún más cuando el objetivo es grabar sesiones de clase completas durante unidades didácticas completas dos veces al año. Además, una vez que se gana el acceso a la escuela, no todos los profesores están listos para o pueden colaborar en el proyecto. La experiencia previa y el conocimiento profesional de los docentes fue otro aspecto que no pudo ser controlado y que pudo haber tenido alguna influencia en los resultados. Tres de los profesores eran bastante cercanos en edad y experiencia, pero el otro tenía mucha más experiencia en la enseñanza que el resto. A pesar de estos problemas, los datos recogidos son muy ricos, ya que consisten en unidades didácticas completas e incluyen diferentes profesores y asignaturas, lo que ha permitido análisis muy detallados y variados.

Aunque el estudio ha identificado las prácticas de interacción de profesores y estudiantes a lo largo de los episodios, habría sido interesante haber explorado con más detalle el desarrollo de técnicas de AfL a lo largo de las lecciones desde el principio hasta el final de cada unidad didáctica. Aunque esto puede hacerse en estudios futuros, la longitud de las unidades registradas fue tan variable (de 1 lección hasta 5 ó 6) que los resultados que hubieran salido de ese tipo de análisis podrían no haber sido fiables. De la misma manera, si los temas de las unidades didácticas en cada materia hubieran sido similares, habría ayudado a que este tipo de análisis fuera más sólido. Sin embargo, tener unidades didácticas completas nos ha permitido disponer de datos amplios y fiables para todos los tipos de análisis discursivos realizados en esta investigación.

También habría sido interesante incluir un análisis detallado de la interacción grupal entre los estudiantes, especialmente en las clases AICLE en las que se implementó AfL, ya que este tipo de interacción ocurre con mayor frecuencia. Como se explicó en el capítulo 4, estos episodios sólo se analizaron más a fondo en términos de patrones IRF siempre que hubo interacción con el profesor sobre el contenido, la lengua o ambos. Sin embargo, es cierto que sería necesario un análisis más detallado de la interacción entre alumnos para completar el análisis completo de la interacción en el aula. Este tipo de discurso no debe ser subestimado o ignorado. Al contrario, su importancia debe enfatizarse (Barnes 1975).

Finalmente, se realizaron entrevistas con los profesores después de cada unidad grabada, ya que son un valioso ejercicio de reflexión para los mismos sobre su práctica. Dichas entrevistas no fueron utilizadas en esta disertación porque el foco estaba en la interacción del aula. Sin embargo, estas entrevistas pueden ser analizadas para futuras investigaciones, a fin de dar una idea de hasta qué punto los profesores ponen en práctica conscientemente estrategias AfL o estrategias que se alejan de esta pedagogía.

2.6 Investigaciones futuras

Con esta investigación como punto de partida, hay muchas otras ideas que se deben considerar para futuras investigaciones. Por ejemplo, el número de asignaturas podría ser ampliado. En esta investigación, hemos tratado cuatro materias diferentes, pero éstas pueden ser aumentadas. Esto podría conducir a la construcción de pedagogías AfL específicas de cada asignatura, es decir, cómo AfL puede ayudar a desarrollar los diferentes géneros y funciones del discurso cognitivo (Dalton-Puffer 2013, 2016) requeridos en diferentes materias en las clases AICLE.

También podría ser interesante ampliar los datos y revelar las diferencias (si las hay) y las similitudes en la implementación de AfL en las clases AICLE, en las clases de contenido en la lengua materna y en las clases de inglés como lengua extranjera. De esta manera, se podría descubrir si las características específicas de diferentes tipos de contextos pueden tener un efecto sobre la interacción AfL y su implementación, con posibles contribuciones para el diseño de pedagogías AfL específicas al contexto.

Aparte de ampliar los datos a más asignaturas y más tipos de contextos, también se podrían compilar datos de más niveles más académicos. El presente estudio se ha centrado en el último ciclo de primaria (quinto y sexto año). Un estudio similar, pero en la enseñanza AICLE en secundaria, sería muy interesante para ver si la implementación de AfL sería diferente en

términos de discurso de clase en este nivel académico, si la co-construcción de AfL tiene las mismas características que en las aulas de primaria, cómo la implementación de AfL en Educación Secundaria afecta el aprendizaje de los estudiantes, etc. En España, sin embargo, es difícil encontrar institutos que implementen AfL. En consecuencia, este estudio sería muy difícil de llevar a cabo.

Además, un análisis cuantitativo detallado de los patrones IRF dentro de los diferentes tipos de episodios destacados en esta investigación proporcionaría una imagen más amplia de la interacción en el aula, lo que contribuiría a hacer sólidas conclusiones sobre si diferentes episodios desencadenan ciertos tipos de preguntas y *feedback* y, consecuentemente, afectan a la implicación del estudiante en tal interacción.

Con el modelo de implementación de AfL propuesto en esta investigación, junto con los resultados de otros estudios actuales y futuros, se podrían poner en marcha programas de formación de profesores para implementar la pedagogía AfL. Después de un tiempo de implementación (1-2 años), se podría llevar a cabo un estudio para averiguar qué efectos positivos ha tenido tanto en profesores como en estudiantes (ver Harrison y Swaffield 2003 en el Proyecto KMOFAP).

Como se señaló anteriormente, en el presente estudio, el aprendizaje como tal no ha sido el enfoque principal. El aprendizaje sólo puede inferirse de la interacción o sólo se puede afirmar que la interacción proporciona oportunidades para el aprendizaje. Para una investigación más profunda, sería interesante añadir pruebas (pre y post) para poder medir de una manera más objetiva lo que los estudiantes han aprendido o no (Ruiz-Primo y Furtak 2006). Sin embargo, este no fue el propósito del presente estudio. El aprendizaje de los estudiantes sobre el contenido y / o la lengua en una clase No-AfL puede ser comparado con el aprendizaje de los estudiantes en una clase AfL y ver si las diferencias son estadísticamente significativas. En caso de que surjan diferencias significativas a favor del grupo AfL, una de las posibles causas puede ser la implementación de AfL.

Como se mencionó en la sección anterior, se realizaron entrevistas con los profesores, aunque no se utilizaron. El análisis futuro de estas entrevistas podría agregar una nueva dimensión al estudio. Como Cowie y Bell (1999) descubrieron, en su estudio los maestros implementaron AfL, pero no eran conscientes de ello. Las entrevistas con los docentes pueden ser una forma útil de estimular una mayor reflexión por su parte, permitiéndoles tomar conciencia de lo que hacen (Cowie y Bell, 1999). Además, la información podría obtenerse no sólo con respecto a la conciencia de los profesores AfL sobre sus técnicas interaccionales de AfL, sino también con

respecto a las prácticas de interacción de los profesores No-AfL, que a veces también revelaban características de AfL.

Del mismo modo, realizar entrevistas con los estudiantes sobre sus impresiones sobre su aprendizaje y motivación sería muy importante. Al final del día, AfL se centra principalmente en los estudiantes y en su aprendizaje, por lo que sus puntos de vista sobre la pedagogía AfL deben ser tan valiosos como los de los profesores. En última instancia, para una investigación más profunda sobre AfL, involucrar a los estudiantes y escucharlos debe ser crucial.

2.7 Observaciones finales

A la luz de esta investigación, se ha concluido que la implementación de AfL en las aulas AICLE puede tener importantes beneficios para el tipo de interacción que se desarrolla dentro de clase, ya que esta interacción combina el aprendizaje de contenidos a través de una lengua extranjera y el aprendizaje de usar dicha lengua extranjera apropiadamente. La interacción resultante en las clases AfL AICLE es más dialógica y receptiva, más alineada con la enseñanza como proceso de investigación que con la enseñanza como proceso de transmisión. Este tipo de interacción es también mejor para la integración del contenido y la lengua y para la evaluación, como ya han subrayado algunos investigadores (Llinares et al., 2012).

Los resultados presentados a lo largo de esta disertación me han llevado a proponer un modelo para implementar AfL en la interacción en el aula AICLE. Este modelo consiste en preguntas de los profesores para que los estudiantes piensen y razonen; en *feedback* ofrecido por los profesores para construir y ampliar las contribuciones de los estudiantes y hacerles reflexionar sobre el aprendizaje; y en episodios que desencadenan la reflexión sobre el proceso de aprendizaje.

Esta investigación tiene la intención de llenar varias lagunas existentes en la investigación sobre AICLE y AfL. Desde la perspectiva de AfL, no hay tantos estudios centrados en cómo se desarrolla esta pedagogía en las aulas, a pesar de su importancia (Black y Wiliam 1998a; Leung 2004; Leung y Davison 2009). Más específicamente en AICLE, AfL apenas ha sido investigado (véase Llinares et al., 2012 para una breve introducción). Una excepción es Basse (2016), quien, utilizando este mismo corpus, investigó AfL en las aulas AICLE en relación con las estrategias de motivación en la lengua extranjera usadas por el profesor y la motivación y habilidades meta-cognitivas del estudiante. Sus resultados mostraron que los profesores de AfL utilizaron estrategias de motivación en la lengua extranjera más frecuentemente y de una manera más variada, lo que dio lugar a un discurso más motivacional. En cuanto a la motivación de los

estudiantes, no hubo diferencias significativas entre las escuelas AfL y No-AfL, aunque los estudiantes con menor rendimiento en los colegios AfL reflexionaban más críticamente que aquéllos en las escuelas No-AfL cuando se evaluaban a sí mismos. La presente investigación es, pues, uno de los primeros estudios que se centran en las prácticas AfL en contextos AICLE, y el primero en centrarse en la interacción del aula AfL AICLE. Desde el punto de vista de AICLE, es muy importante que este modelo tenga estudios sobre la enseñanza y el aprendizaje reales en las aulas, como han afirmado recientemente Cenoz, Genesee y Gorter (2014). Por último, la investigación actual contribuye a llenar el vacío existente en los estudios AICLE en el nivel de educación primaria y, más específicamente, la interacción que tiene lugar en las aulas AICLE de primaria.

Como conclusión, el presente estudio ha sido una contribución tanto a la investigación como a la práctica de AICLE y AfL, adoptando una perspectiva discursiva. A través de un análisis en profundidad de la interacción en el aula construida conjuntamente por el profesor y los estudiantes, se han destacado las principales características del discurso AfL en las clases AICLE en contraposición al discurso No-AfL en estas clases. Los tipos de preguntas y *feedback* utilizados por los docentes afectan los tipos de respuestas dadas por los estudiantes, así como la posibilidad de volver a tener la palabra e incorporar dicho *feedback* por parte de los estudiantes. El uso de ciertos tipos de preguntas y *feedback* y no otros ayuda a construir un tipo de interacción dinámica, contingente y con capacidad de reacción. Esta investigación también ha abordado la integración del contenido y la lengua extranjera (elemento que hace a las aulas AICLE diferentes), y cómo la aplicación de AfL puede ser útil en este asunto. Por último, esta investigación ha contribuido a demostrar cómo diferentes asignaturas pueden requerir diferentes tipos de discurso o géneros.

Assessment for Learning in Primary CLIL Classrooms and its Co-Construction in Classroom Discourse

1

Introduction

1.1 PURPOSE AND SCOPE OF THE STUDY

The present dissertation addresses the role of *Assessment for Learning* (AfL) or Formative Assessment in primary *Content and Language Integrated Learning* (CLIL) classrooms in the community of Madrid. The main purpose of the study is to characterize how discourse is jointly co-constructed by teachers and students in whole-class interactions in the CLIL/bilingual education programmes under analysis, and to explain how classroom discourse is aligned (or not) with an AfL pedagogy, and its effects on students' language and content use.

Content and Language Integrated Learning (CLIL) refers to "any educational situation in which an additional language and therefore not the most widely used language of the environment is used for the teaching and learning of subjects other than the language itself" (Wolff 2007:16). CLIL has been supported during almost three decades by many European projects and it was conceived to create multilingual citizens in a multilingual Europe (Whittaker & Llinares 2009; Dalton-Puffer 2011). This teaching method, which, as the name states, consists in the integrated teaching of content subjects and a second/foreign language, provides students with extra time of exposure to the foreign language and practice in it (Dale & Tanner 2012), but it also entails an integrated approach to learning language and content (e.g. Llinares 2015). Many different European countries (over 80%) are implementing CLIL in one way or another (Dale & Tanner 2012). Spain is one of the countries in which CLIL has been implemented very rapidly (Llinares & Dafouz 2010). In Madrid, in primary education, there are two CLIL projects taking place: the Ministry of Education/British Council Project and the Community of Madrid project (Llinares & Dafouz 2010). In this thesis, data come from schools participating in both projects, the focus of the study being the comparison between those that implement Assessment for Learning and those which do not.

As a result of the fast implementation of CLIL in Europe, research on CLIL has been extensive, both at the empirical and theoretical levels (Dalton-Puffer 2011). Many aspects have been investigated at the empirical level (see Dalton-Puffer 2011 for a review), but it seems that assessment is one of the areas in CLIL research that has hardly been explored, especially Assessment for Learning (but see Basse 2016; Llinares, Morton & Whittaker 2012; Pascual & Basse 2017).

Assessment for Learning had its turning point with Black and Wiliam's (1998a) review of the literature on this type of assessment, in which the authors highlighted the benefits of this type of methodology, and which was followed by other studies and – even – some political changes. However, the studies reviewed in Black and William's extensive literature review mainly focused on mainstream classrooms (Rea-Dickins 2008). Exceptions are studies like Leung and Mohan (2004), Rea-Dickins (2001), Rea-Dickins and Gardner (2000), and Edelenbos and Kubanek-German (2004), based on *English as an Additional Language* (EAL), *English as a Second Language* (ESL), and *English as a Foreign Language* (EFL) primary classrooms. AfL in CLIL classrooms is still an under-researched issue to which this dissertation aims to contribute.

This study aims to conduct a thorough and detailed analysis of how AfL is constructed in primary CLIL classrooms through teacher-student interactions across a variety of subjects. This research is innovative, as AfL theory has only been applied to CLIL classrooms in one previous study on motivation and assessment for learning using the same contexts studied here (Basse 2016); and hence more research in this area, and on classroom discourse in particular, is clearly needed (Nikula, Dalton-Puffer & Llinares 2013). It is necessary to study AfL in CLIL classrooms not only as a form of assessment, but also to find out whether AfL discourse can have positive effects on the quality of the interaction. This may have important implications for CLIL pedagogy as high-quality interaction in CLIL classrooms has been seen to positively affect students' content and language engagement (Dalton-Puffer 2009; Nikula, Dalton-Puffer & Llinares 2013). Thus, one of the objectives of this thesis is to recommend a model for characterising AfL discourse, which could be used as the basis for teacher training courses and the creation of materials for the implementation of AfL in CLIL classrooms.

Assessment for Learning theory claims that teaching, learning, and assessment are part of the same process (Black & Wiliam 1998a, b). It is a theory that is closely related to sociocultural theory, in that interaction is considered to be fundamental for learning (Vygotsky 1978; Van Lier 1996; Lantolf & Thorne 2006). As Leung and Mohan (2004: 336) state, "much of the formative work is interactionally realized through teacher-student talk". If interaction is

essential for learning, and if learning, teaching, and assessment are the same entity, then it follows that interaction is also essential for Assessment for Learning (Black & Wiliam 1998b). In the same vein, Leung (2004: 29) expresses the necessity of paying attention to classroom interaction in order to study and understand assessment for learning: "the socially co-constructed nature of formative teacher assessment therefore makes it necessary to attend to classroom interaction and classroom discourse as a key site for empirical investigation" (see also Black & Wiliam 1998a; Edelenbos & Kubanek-German 2004). Despite the reported relevance of interaction in Assessment for Learning, there exist very few studies analysing actual interaction in relation to AfL (Leung & Mohan 2004; Harlen & Winter 2004; Rea-Dickins 2001; Gardner & Rea-Dickins 2002; Anderson et al. 2007; Ruiz-Primo & Furtak 2006, 2007). In that respect, this thesis would contribute to this area in AfL research in a new context (i.e. CLIL).

In brief, by analysing classroom interaction and relating it to Assessment for Learning theory in CLIL classrooms, this dissertation aims to fill two important research gaps: one, to characterize the discourse of AfL in CLIL or bilingual education programmes, and two, to explore the effects of AfL interaction on students' content and language use and learning. Through the implementation of the AfL model for CLIL classrooms proposed in this dissertation, CLIL teachers can interactionally react to students' content and language misunderstandings by integrating the two aspects or by paying special attention to one or the other. Feedback intended to focus on content may also be beneficial for students' linguistic competence, and feedback focusing on language can also be very valuable for students to meet content learning goals (Llinares et al. 2012).

1.2 CLIL

CLIL has often been defined both as a type of programme and as a type of methodology. This dichotomy is reflected in the macro and micro levels of CLIL research (Dalton-Puffer & Smit 2007; see also Leung 2005a for bilingual education in general): the micro level focuses on the participants, results, and processes in CLIL methodological implementations; the macro level is concerned with the characteristics of CLIL programmes. Dalton-Puffer and Smit (2007) emphasize the importance of the micro level observing that, without it, the macro level would not exist. This dissertation is framed at the micro level, and it will contribute to this growing stream of research in CLIL (see, for example, Llinares & Whittaker 2009; Dalton-Puffer 2007; Dalton-Puffer & Nikula 2006).

It has been claimed that one of the aspects that distinguish CLIL from other types of bilingual education (such as immersion or content-based instruction), is the fact that the language of instruction is a foreign language and not a second or indigenous language. Furthermore, the fact that this foreign language is mostly English (as in the present study) led Dalton-Puffer et al. (2010) to coin the term CEIL (Content and English Integrated Learning). Another distinctive aspect is, as the term CLIL indicates, the integration of both language and content objectives (Dalton-Puffer & Smit 2007). However, according to Dalton-Puffer and Smit (2007), in Europe most of the curricula are content-based, meaning that the language aspects are often left aside. This could be explained by the fact that European governments do not invest enough resources in CLIL teacher training and CLIL implementation (Dalton-Puffer 2011). In Spain, pre-service training programmes for CLIL are almost non-existent, and the in-service programmes that exist are still scarce (Fernández Fontecha 2009) and rarely address the role of language in content teaching.

Language learning goals should not only remain visible in CLIL curricula, but need to be made explicit for both teachers and students (Mohan et al. 2010; Llinares, Morton & Whittaker 2012). Although language is a key component in any type of assessment, AfL becomes especially relevant for CLIL because it is both planned and reactive (Llinares et al. 2012). When learning objectives are being planned, not only concepts and skills need to be taken into consideration, but also the language that goes hand in hand with them (see also Coyle et al. 2010). On the other hand, CLIL teachers also need to be sensitive to their students' needs, and consequently they need to react and adjust their teaching accordingly, focusing on the language necessary to acquire different content goals. The taxonomy presented in this study accounts for possible distinctions between attention to language or attention to content. Although it can be argued that content and language cannot be separated from one another (Halliday 1978; Widdowson 1978; Coyle et al. 2010), this distinction has been made in order to highlight the duality of CLIL (content and language). Precisely out of the CLIL duality emerges one of its main challenges: learning a foreign language and using that language to understand, express, and learn content appropriately in teaching/learning scenarios where the cognitive and linguistic levels of students can vary substantially (Coyle et al. 2010).

1.3 THE THEORETICAL BACKGROUND: ASSESSMENT FOR LEARNING AND CLASSROOM DISCOURSE

The theoretical perspectives applied in this dissertation are two: Assessment for Learning and classroom discourse.

There has been much recent discussion on classroom-based teacher assessment. It has recently been argued that assessment should be integrated with teaching, be formative, and for-learning (Leung & Mohan 2004). This type of assessment has sometimes been seen as a more desirable alternative to summative assessment, while others view it as a complementary practice (Leung 2004). Assessment for Learning, also called formative assessment or classroom-based teacher assessment, contrasts with summative assessment or formal testing, as the latter is centred on the learning product (Spolsky 1992) as opposed to the former, which focuses on the learning process (Leung & Mohan 2004; Black & William 1998a; Rea-Dickins 2001). Assessment for Learning has been defined as a form of assessment whose ultimate goal is improving teaching practices and learning processes and outcomes (see e.g. Shohamy 1992; Harlen 2005; Leung & Mohan 2004; Black & William 1998a, b). As some researchers point out (Leung 2004; Leung & Mohan 2004), there is a need to examine formative teacher assessment in depth, just as standardized assessment has been, if we want to understand how this type of assessment is really accomplished in classroom interaction, and if we want to appropriate both theory and research methods "in the study of this highly complex and dynamic aspect of teaching-learning interface" (Leung & Mohan 2004: 338). If formative assessment is realized through teacher-student interaction, students' assessment necessarily has to be reconceptualized as discourse: "formative assessment has to take account of the interactive and contingent nature of student performance in the classroom which is dynamic and co-produced with the teacher and others" (Leung 2004: 22).

Assessment for learning is meant to be integrated into everyday teaching and learning activities and teachers are supposed to be able to respond to the contingent needs of students in the learning process (Leung 2004). That is why formative assessment cannot be easily accommodated within a set of pre-specified criteria (Leung 2004). The QCA (Qualifications and Curriculum Authority),,a quasi-official government agency in England, presents the principles of Assessment for Learning as follows:

> Assessment for Learning happens all the time in the classroom. […] a pupil needs to know where he or she is and understand not only where he or she wants to be but also how to "fill the gap". This involves both teacher and the pupil in a process of continual reflection and review about progress. When teachers and peers provide quality feedback,

> pupils are empowered to take the appropriate action. Teachers adjust their plans in response to formative assessment (in Leung 2004: 22).

As Clarke suggests (1998: 117), "[f]ormal testing, while useful for some purposes, does not in itself appear to raise standards. Formative assessment strategies can". In the same way, Black and Wiliam's work (1998a: 3) showed that "innovations which include strengthening the practice of formative assessment produce significant, and often, substantial, learning gains. These studies range over ages (from five-year-olds to university graduates), across several subjects, and over several countries".

In AfL, classroom interaction analyses are important for two main reasons: first, because it is in interaction that assessment for learning takes place (Leung 2004; Leung & Mohan 2004; Rea-Dickins 2001; Black & Wiliam 1998b); and second, because if we get to know how this type of assessment operates, its implementation could be more effective. If we understand that learning is jointly co-constructed by individuals (teacher and students in the case of an educational context) (Vygotsky 1978; Hammond & Gibbons 2005), and that one of the crucial aspects for learning is interaction (Van Lier 1988, 1996; Mortimer & Scott 2003; Barnes 1975; Hall & Walsh 2002; Gibbons 2003; Vygotsky 1978; Lantolf & Thorne 2006; Llinares & Whittaker 2009), then interaction is also fundamental for AfL, as anything that can improve or enhance learning is important for AfL (Black & Wiliam 1998b; Leung & Mohan in press; Leung 2004; Rea-Dickins 2001). As Leung and Mohan point out (2004: 336), "much of the formative work is interactionally realized through teacher-student talk". As assessment and learning are interactive and socially constructed, classroom discourse and interaction will reflect and construct formative assessment. As a consequence of this tight relationship, one could conclude that classroom interaction and classroom discourse need further investigation to understand formative assessment and its socially co-constructed nature (Leung 2004: 29). The importance of discourse is highlighted not only in AfL but also in sociocultural theory and other approaches to second language acquisition in the classroom. One of the reasons why this thesis can be a valuable contribution both to AfL and CLIL research is its detailed analysis of classroom discourse in CLIL Primary classrooms and its relationship to AfL pedagogy, always directed towards improving learning.

One of the classroom discourse patterns that has been most widely studied in relation with the opportunities it offers for language use and learning is the *Initiation-Response-Feedback exchange* (IRF) (Sinclair & Coulthard 1975). This pattern has been widely criticized on the basis that teachers do most of the talking and students are not allowed to pursue their own topics or ideas (Barnes 1975; Lemke 1990; Wells 1993; Mortimer & Scott 2003). In contrast, in other

studies, this pattern has been said to be effective for certain purposes, especially for checking students' understandings and guiding their learning (Mercer 1992; Seedhouse 1997; Nassaji & Wells 2000). The present study will explore the role of IRF patterns in AfL in CLIL classes.

Within the IRF exchange, teachers' questions have been extensively studied from different perspectives in a variety of class types. These perspectives include the *open/closed* formal dichotomy (Barnes 1969; Musumeci 1996), the *display/referential* functional dichotomy (Long & Sato 1983; Romero & Llinares 2001), the *content* perspective (Dalton-Puffer 2007; Llinares & Pascual 2014), and the *cognitive demand* approach (Bloom et al. 1956; Redfield & Rousseau 1981). Teacher questioning is the key for successful teaching (Wragg & Brown 2001), and therefore one of the most important characteristics of AfL, since AfL is committed to successful teaching and learning (Black et al. 2004; Black et al. 2003; Wiliam et al. 2004; Black & Wiliam 1998b). Several studies have claimed that teachers' questions should encourage discussion, reveal students' understandings, and require long responses (Harrison & Howard 2009). In this thesis, with the help of a specific taxonomy, both CLIL teachers' questions and the corresponding students' responses are analysed, as students' contributions are also part of classroom discourse, and they are crucial for teachers to guide their learning (Van Lier 1996).

The third part of the IRF exchange (follow up/feedback) has also been the object of much investigation due to its importance for quality interaction (Lee 2007; Lyster 2007; Alexander 2004; Nassaji & Wells 2000). Some researchers distinguish between IR*F* and IR*E* (Initiation-Response-*Evaluation*), depending on whether the teacher offers *evaluative* or *non-evaluative* feedback (Mehan 1979; Van Lier 1996; Hall & Walsh 2002; Wells 1993). A basic characteristic of AfL is that feedback should be helpful for guiding and improving, and needs to aim at moving learning forward (Leung 2007; Harrison & Howard 2009; Heritage 2010; Leahy et al. 2005). One of the ways in which this can be achieved is by using students' responses to engage them into discourse (Black & Wiliam 2009). If the feedback move is used not only to assess but also to facilitate students' learning, then meaning can be co-constructed and learning opportunities can be greater (Wells 1993). This study hopes to illustrate a range of different functions for which teachers may use the third turn, and to explain the different implications that each of them has for the quality of interaction and learning opportunities.

In CLIL classroom discourse, interaction is an opportunity for both language development and content learning. Thus, if students are engaged in interaction, foreign language and content knowledge can develop (Morton 2012). As in EFL, it seems that IRF sequences are very frequent in CLIL classrooms (Dalton-Puffer 2007). However, when compared to EFL lessons, IRFs in CLIL seem to give space to more dialogic teaching, with longer student responses and a

variety of functions performed in the third move (Nikula 2007a). The third turn is crucial in any type of classroom, but even more in the case of CLIL: teachers need a variety of interactional strategies to help students understand and learn concepts which are co-constructed in a foreign language (Evnitskaya 2012). In this study, the third turn is shown to be helpful for teachers to assess students' content and language knowledge, and to guide students' content and language learning.

Finally, in classroom discourse, long exchanges of IRFs give rise to distinct episodes. During lessons, then, different episodes emerge, as many studies have already highlighted and analysed (Frölich, Spada & Allen 1985; Bloome et al. 2009; Snell & Lefstein 2011; Berg 2009). As posited in this dissertation, different types of episodes can be more or less aligned with an AfL pedagogy, depending on what is being done and what the goal of the episode is. If the type of episode is differentially aligned with AfL, then it also has a differential impact on the type of questions and feedback used by the teacher. Likewise, the type of episode can also affect how engaged students are in discourse.

1.4 THE STUDY: OBJECTIVES, RESEARCH QUESTIONS AND HYPOTHESES

As pointed out above, assessment in CLIL is one of the areas which needs further research, especially when it comes to the relationship between formative assessment and CLIL (Llinares, Morton & Whittaker 2012). As much of the formative work is realized through teacher-student talk (Leung & Mohan 2004), the main objective I intend to achieve in this thesis is to offer a detailed account of how teachers construct Assessment for Learning in CLIL classrooms and how that, in turn, affects students' participation; in other words, the purpose is to find out how discourse shapes AfL through various interactional patterns and teacher strategies. The specific research questions related to this main objective are the following:

1. What are the interactional features and strategies which characterize AfL discourse in Primary CLIL classrooms?
2. Are these patterns specific for AfL classes or are they also found in similar classes where AfL is not implemented?
3. Are there differences in these interactional features and strategies that characterize AfL discourse across subjects? Are there differences in the discourse of the same teacher across different subjects?
4. To what extent do AfL strategies affect students' participation in classroom discourse?

The objectives of this investigation are:

1. Characterize how AfL discourse is constructed at different Primary school levels, across different subjects and with different teachers.
2. Compare the discursive characteristics of AfL in schools which implement this approach and the discourse patterns used in similar CLIL schools that do not specifically implement it.
3. Compare teachers' interactional patterns across subjects.
4. Discover how AfL strategies affect students' participation in the discourse.

In light of the objectives and research questions expressed above, the corresponding hypotheses that I will be working with are the following:

1. The first hypothesis is that AfL in CLIL is constructed through a specific type of discourse, with the use of certain types of questions, certain episodes and certain types of feedback by the teacher.
2. The second hypothesis articulates that there will be differences in the discourse patterns generated by teachers in AfL schools and teachers in Non-AfL schools.

The first two hypotheses are based on AfL theory, which highlights the importance of good teacher questioning, reducing factual questions and using high order questioning, and the importance of quality teacher feedback, which moves away from mere evaluation and tries to scaffold students' learning processes (Black & Wiliam 1998a, b; Black et al. 2003; Harrison & Howard 2009).

3. The third hypothesis states that a) there will be differences in the interactional patterns used in different subjects and b) differences will be found in the discourse of the same teacher teaching different subjects.

This is based on the evidence that AfL varies depending on different factors, one of them being that the same teacher can invoke different pedagogic assumptions and principles in different contexts (subjects) and with different students (Black et al. 2004; Wiliam 2006; Black & Wiliam 1998a; Black & Wiliam 2009; Hodgen & Marshall 2005; Leung 2004; Torrance & Pryor 1998).

4. The fourth hypothesis claims that AfL strategies will influence positively on students' participation and contributions. As a consequence, it is hypothesized that there will be differences in the types of responses given by students in AfL schools and Non-AfL schools.

1.5 THE STUDY: DATA AND METHODOLOGY

Formative teacher assessment has been supported by policy makers in diverse countries such as England, Australia, and Hong Kong. That is not the case in Spain, where only the schools which are part of the British/MEC project consciously implement this type of assessment. Half of the schools in which the present study was carried out belong to this project, and half belong to the Comunidad de Madrid project where Assessment for Learning is not being implemented. The fact that the data come from CLIL primary schools is an important contribution of this thesis, since most CLIL research on interaction (and CLIL research in general as well) has focused on secondary education (Dalton-Puffer 2007; Llinares et al. 2012; Llinares & Whittaker 2009; Nikula 2005, 2008; Moore 2011), but fewer studies centre on tertiary (Dafouz Milne & Núñez Perucha 2010; Smit 2010) or primary levels (Llinares & Lyster 2014; Pastrana 2010; Serra 2007).

The data for the present study include classroom sessions from four schools. In each school, sessions on two subjects taught by the same teacher were recorded. The school levels are grades 5 and 6 (students are 10-11 and 11-12 years old, respectively) and the subjects recorded are Science, Arts, Drama, and Citizenship. There have been two rounds of recordings in each school in the two different subjects, one at the beginning of the school year, the other one at the end. In this way, data were more reliable because the variables of topic of the unit and recording time could be controlled, avoiding distortion of results because students liked or did not like one specific topic, or because students were more or less familiar with their teacher, his/her teaching style, and their classmates. Likewise, differences (if any) between the co-construction of discourse at these two different moments throughout the year could be detected. Complete didactic units in each subject were recorded, which means that there are two complete didactic units for each of the disciplines recorded in each school. Didactic units varied from two to five classroom sessions. All in all, 44 classroom sessions are used for this thesis, which amounts to a total number of approximately 50 hours and a total number of around 350,000 – 400,000 words.

All teacher-student interactions were analysed, as whenever there is interaction, there is a chance that AfL is implemented. Each classroom session was divided into *episodes*, "which consists of all the talk produced in carrying out a single activity or one of its constituent tasks" (Nassaji & Wells 2000: 383). These were classified according to the purpose of the activity; in other words, what the teacher and students were doing in them.

In this study, the exchange is the main unit of analysis and its different constituents (Initiation-Response-Feedback) were analysed as follows:

1. *Initiations*: These were classified according to who initiates: teacher or student. When the teacher initiates and that initiation is a question, this was classified following Dalton-Puffer's (2007) typology for CLIL classrooms: questions for facts, questions for explanations, questions for reasons, questions for opinions, and meta-cognitive questions (see full classification in chapter 4). Student initiations were also scrutinized, paying special attention to the type of contribution they were making: asking a question, expressing personal opinion, stating a fact, explaining, arguing, etc. This could provide information as whether the type of teaching was more dialogic or authoritative and how this correlates or not with the essence of formative assessment.
2. *Responses*: As a way of measuring length and complexity in the students' responses, these were divided into minimal responses, truncated responses, and T-unit responses (further divided into one-phrase, one-clause and more than one clause). Responses were analysed and compared taking into account the type of class (AfL or not), the type of question previously asked by the teacher, and the subject taught.
3. *Follow-up/feedback moves*: Teachers' feedback moves were also classified as feedback, which is another key element in AfL theory (see for example Black & Wiliam 1998a; Torrance & Pryor 2001; Wiliam et al. 2004; Sadler 1998). The aim of this research was to create a taxonomy that described the different and various types of feedback that primary CLIL teachers use, with special attention to those types that clearly help to construct AfL.

1.6 OVERVIEW OF THE THESIS

The present dissertation is divided into nine chapters. The first chapter (the present one) is the introduction, concerned with establishing the purpose and scope of the study, as well as the objectives, research questions, and hypotheses.

The second chapter is entirely devoted to CLIL as an educational model. Several aspects of this model are explored, such as what it is, its precursor methodologies, the situation of CLIL in Spain and more specifically in Madrid, CLIL methodology, lines of research, and learning outcomes.

Chapter three deals with the theoretical framework necessary to understand, interpret, and carry out this research. As pointed out earlier, there are two main strands in this section: classroom discourse and assessment for learning. Within classroom discourse, IRF patterns are the key focus,

as well as discourse and interaction in CLIL. In addition, the effects of different types of feedback on students' language learning will be explored. In the AfL section, AfL and its characteristics are explained in a detailed way, especially concentrating on feedback and self- and peer-assessment. Finally, the role of classroom discourse in assessment for learning is explored.

The fourth chapter presents all the data that have been used for this study, as well as how the data have been treated and analysed.

Chapters five, six, seven, and eight focus on the results. In all of them, examples from the corpus are included in order to illustrate the results obtained from the analysis. Chapter five focuses on the results obtained from the episode analysis. Chapter six concentrates on the analysis of teachers' types of questions, starting in this way the analysis of IRF sequences. To continue with the analysis of IRFs, students' responses are then described and discussed in chapter seven. In the last place, chapter eight focuses on teachers' feedback.

Although every chapter on results includes a discussion section, chapter nine is devoted to discussing the results at a more global level: the main findings of the study will be highlighted in the light of the initial objectives set for the study. Also, the role of interaction in AfL CLIL classrooms will be discussed. This chapter also provides a proposal for a model of AfL discourse practices in CLIL classrooms which will enable teachers to assess students' content and language learning gaps and adjust their teaching in order to close those gaps and thus improve students' learning. Pedagogical applications, limitations of the study and further research are also dealt with in this final chapter. Finally, some concluding remarks round up all the work presented throughout this thesis.

1.7 CHAPTER SUMMARY

The present chapter is a guide to what can be found in this dissertation. Firstly, an overview of the purpose and the scope of the study has been offered, presenting the different frameworks and why this thesis might be interesting and contribute with new findings. Secondly, the different theoretical perspectives used for this study have been briefly presented and connected and some important issues have been raised (see chapters 2 and 3 for the extended theoretical framework). Thirdly, the objectives, research questions and hypotheses of the dissertation have been formulated. Next, the data and the methodology used have been concisely explained (see chapter 4 for a detailed account). Finally, this chapter has presented the organization of the thesis, providing a short description of what is to be found in each of its different sections.

2

Content and Language Integrated Learning (CLIL): an educational model

One possible definition of bilingual education can be the use of more than one language for teaching and learning school subjects. However, as simple as it may seem, this educational model varies in different contexts with respect to purposes and curriculum configurations (Leung 2005a: 239). Some experts consider Content-Based Instruction (CBI) or Content-based Learning (CBL) as good umbrella terms encompassing different types of bilingual education programmes (Lyster & Ballinger 2011). This term also entails positive methodological connotations. Cummins (2004:108) refers to CBI as a methodology that "encourages active student participation and learning rather than the typical passive role that has been assigned to students". If one considers CBI an umbrella term, then Canadian French Immersion programmes (FI) or Content and Language Integrated Learning (CLIL) in Europe would represent types of CBI. However, CLIL could also be used as an umbrella term if we follow definitions such as Wolff's (2007: 16):

> a generic term and refers to any educational situation in which an additional language and therefore not the most widely used language of the environment is used for the teaching and learning of subjects other than the language itself.

CLIL, then, covers a wide range of practices in which curricular content is taught through the medium of a language that is not the learners' L1. This foreign language is, in most of the cases throughout Europe, English.

This chapter has the overall purpose of situating CLIL as an educational practice. To do that, some considerations on other non-European content-based programmes will be described (namely Canadian French immersion programmes and CBI programmes in the US), as they are

thought to be CLIL's precursors. Next, the chapter focuses on the situation of CLIL in Europe, Spain, and Madrid, where the present study is set. This is followed by a thorough section on CLIL methodology, in which matters such as CLIL benefits, CLIL challenges, and the integration of content and language are addressed. Potential similarities and differences between CLIL and other content-based programmes are then pointed out. The focus of the chapter next shifts to CLIL research, explaining the currently existing lines of research. Finally, the last section of the chapter is devoted to describing the linguistic features characteristic of different subject disciplines present in this study.

2.1 BILINGUAL EDUCATION: NON-EUROPEAN CLIL-RELATED MODELS

As specified above, CBI or CBL used as generic terms encompass different types of bilingual education programmes. Lyster (2007) uses the term content-based instruction as an umbrella term to refer to classes where subject matter offers opportunities to second language learners of processing and negotiating the target language (TL). Since content and language cannot be separated (Halliday 1978) and since language and cognitive development go hand in hand (Snow, Met & Genesee 1989), then content-based instruction provides the cognitive basis for language learning and the motivation of purposeful communication (Lyster 2007).

CBI programmes range from content-driven (i.e. immersion) to language-driven programmes (Met 1998). The latter consist of language classes in which thematic or content knowledge is used in order to practice and learn the language. In content-driven programmes, literacy and language development are promoted through subject content learning, and both language and content knowledge are assessed (Lyster & Ballinger 2011). On the contrary, in language-driven programmes, content knowledge is not the object of assessment (Lyster & Ballinger 2011). In the middle of the continuum, there are those programmes which teach one or two content subjects in the target language, along with foreign language (FL) or language arts lessons (Lyster & Ballinger 2011). This latter description could well be applied to many CLIL programmes in Europe, including Spain.

Since some of these types of CBL can be said to be forerunners of CLIL, the following subsections will be devoted to describing two influential models in CBL/CBI: Immersion programmes in Canada, and CBI in the US.

2.1.1 Canadian Immersion Programmes

Immersion programmes were first developed in Canada in the 60s. In them, French was taught through content subjects with the aim of providing the English-speaking population with an opportunity to learn French, the other official language of the country (Lyster 2007).

Different types of immersion programmes were set up, varying in two aspects: starting age/grade and amount of instruction (Genesee 2004; Lyster 2007). Regarding the first aspect, immersion programmes are divided into early immersion programmes, which begin in grade 1 and where literacy in the second language is taught before literacy in the mother tongue (which is delayed until grades 2–4); middle immersion programmes, starting around grade 3; and late or delayed immersion programmes, which begin in grade 6 or even high school. Regarding the second criterion, there are total and partial immersion programmes: the former with all the instruction except English language arts taught through French, the latter with varying percentages, but at least 50% of the curriculum taught through the second language (Lyster 2007).

Research on French immersion programmes revealed some positive findings related to students' motivation and competence in the second language: learners were more motivated and students reached higher levels of proficiency in the second language than non-immersion students, especially when it came to listening and reading comprehension and fluency (Lyster 2007: 14; Harley et al. 1990; Swain & Lapkin 1982). However, some negative findings were also pointed out. Lyster (1987, 2007) discovered that students immersed in bilingual programmes could communicate in an efficient way, yet they did not show native fluency or grammar accuracy (see also Swain 2000; Harley et al. 1990; Swain & Lapkin 1990). Gaps in students' grammatical and lexical development were also found (Lyster & Ranta 1997; Lyster 2007). Immersion students were found to acquire an academic register of the TL, but without colloquial lexical variants, which would make communication among peers more authentic (Lyster 2007; Tarone & Swain 1995). In Cummins' (1979) terms, immersion students developed CALP (Cognitive Academic Language Proficiency) before BICS (Basic Interpersonal Communication Skills), contrary to what would naturally happen in the acquisition of the L1.

Regarding the results obtained in the different types of immersion programmes described above, Swain and Johnson (1996) summarize them as follows:

- Early total immersion students performed as well as non-immersion students in content tests. However, early partial immersion learners did not.

- Early total immersion students performed as well as non-immersion students or even better in all aspects of English language skills (after two or three years of delay, given the fact that they first acquired literacy in the L2) and in general academic achievement.
- Early and late immersion students had similar levels in writing skills in French, but were behind native speakers' level. In general, the main shortcomings found had to do with grammatical competence and vocabulary knowledge, rather than discourse aspects of performance. Speaking was the weakest skill.

The results obtained from research in Immersion Programmes led to the conclusion that more emphasis on form and on how form can make meaning is needed in these immersion programmes if students are to reach a native level of grammatical accuracy (Swain 2000). In other words, subject matter teaching is not enough for language learning: the language used to express content needs to be stressed in ways that makes linguistic features salient for L2 learners (Harley et al. 1990; see also Lyster 2007, Gibbons 1998).

While some researchers have said that language is a secondary goal (Genesee 1994), many others have claimed that language and content should share an equal status of importance (Lyster 2007; Met 1998; Allen et al. 1990). Since initial conceptualizations of immersion programmes undermined the extent to which the target language needed to be attended to, Lyster (2007) advocates for a counterbalanced approach (which he refers to as the systematic integration of content-based and form-focused instruction). This approach suggests that the effort required for students to attend to form in meaning-oriented classrooms will influence their interlanguage system (IL) (Lyster 2007: 4). The counterbalanced approach gives students many opportunities of processing and negotiating language through the curriculum, exploiting to the fullest the potential that language has as a powerful cognitive tool for learning (Lyster 2007: 138). Research shows that if content-based teaching increases the demands on students' language systems by emphasising language as well as content, it has positive effects on second language learning outcomes (Netten 1991; Day & Shapson 1996; Genesee 1987). However, in practice, counterbalance is not properly achieved most of the times in Immersion programmes: either insufficient attention is paid to language, or content is left aside (Lyster 2007: 25; Swain 1996). In immersion and content-based classrooms, form and function have traditionally been kept separately, and incidental language gains were believed to occur without making form/meaning relationships explicit, while the content was the primary priority (Lyster 2007: 27; Netten 1991; Lyster 1998d). However, incidental approaches to language learning do not seem to ensure the learning of less salient yet crucial morphosyntactic features of the TL, and their effect on learners' IL development is disappointing (Lyster 2007; Day & Shapson 1996;

Lyster & Mori 2006; Lyster & Ranta 1997). In turn, isolated grammar lessons may have minimal effects on content-based classrooms because learners are exposed to language instruction as something distinct from language use (Lightbown 1998; Lyster 2007). In this way, students will have difficulties to transfer what they learn in language instruction and apply it to language use. Grammar lessons may be remembered in similar contexts but hard to retrieve in communicative interaction (Lyster 2007). On the contrary, language features which are noticed during interaction will be more easily retrieved in communicative contexts (Lyster 2007). In sum, form-focused instruction is believed to be effective when it is done in communicative contexts, since students will be able to transfer it to communicative interactions outside the classroom (Lyster 2007).

In spite of the general support for the counterbalanced approach by the research community, some questions are open to debate, such as the different effects of this model depending on the degree of explicitness; what features can benefit the most from this technique; or to what extent (and how) form-focused instruction has to be integrated in communicative activities (Ellis 2002; Lightbown 1998; Doughty & Williams 1998; Long & Robinson 1998; Lightbown & Spada 2006).

One proposal to focus on form in content classrooms is through proactive and reactive form-focused instruction (Lyster 2007: 44, 2004a, b). Proactive approaches are pre-planned instruction designed for students to notice and practice certain language features that would otherwise pass unperceived. Reactive form-focused approaches are those in which drawing learners' attention to language is done in a spontaneous and unplanned way (Lyster 2007). It is precisely in this reactive way when focus on form may be more effective (better than waiting for the language arts lesson) (Lyster 2007: 47; Long 1991; Lyster 1998c, d). The present study will also analyse CLIL teachers' reactive form-focused approaches.

2.1.2 CBI programmes in the US

CBI in the US is another form of bilingual education that, together with Immersion programmes, has inspired CLIL programmes in Europe. The main difference between the two is that, whereas in Immersion programmes another national language is taught and learnt, CBI is mainly concerned with the teaching of English to US immigrants. In US CBI, students are normally from minority language groups and they have to learn the new language to integrate in society.

In the US, there are three main types of CBI programmes (Lyster & Ballinger 2011): *two-way immersion* programmes, in which a similar number of students having two different mother

tongues are integrated in the same classroom and so curricular instruction is provided in both languages; *content-based ESL*, in which students' proficiency in English is developed through the incorporation of content from subject areas; and *sheltered instruction*, which consists of content courses for ESL learners which are taught by content specialists rather than ESL ones. The latter two, which are common for newcomers to the US, share the fact that students do not receive any L1 support (Lyster & Ballinger 2011). CBI in the US has also come to refer to approaches in the classroom in which the native languages of the learners are used for instruction. Although controversial, they have nonetheless proven to be as effective as non-bilingual programmes (Navés 2009).

2.2 CLIL IN EUROPE, SPAIN, AND MADRID

CLIL has been considered a different way of implementing bilingual education (Mehisto et al. 2008). However, there have been recent debates regarding the specificities of CLIL when compared to other types of bilingual education programmes (see Cenoz et al. 2014; Dalton-Puffer et al. 2014). Cenoz et al. (2014) argue against the uniqueness of CLIL from a pedagogical viewpoint, although they recognize its European label and its uniqueness from a historical point of view. They argue that there is a lack of precision in the definition of the concept, thus making it difficult "to identify features that are uniquely characteristic of CLIL" (Cenoz et al. 2014: 13). As a consequence, and like CBI, their proposal is to conceptualize CLIL as an umbrella term that includes different programme alternatives and learning opportunities (Cenoz et al. 2014). Dalton-Puffer et al. (2014) respond to Cenoz et al. (2014) emphasising the European background of CLIL, the fact that the language of instruction is mainly a foreign language and the important focus given to *integration* in CLIL research. However, in spite of research interests and although the term CLIL itself entails content and language *integration*, the reality is that most CLIL programmes are content-based, and the existing national curricula do not always contemplate the language requirements of different academic subjects (Dalton-Puffer & Smit 2007).

CLIL has spread globally at a very rapid pace (Dalton-Puffer 2011). The health of CLIL in Europe is very good, especially in some countries like Austria, Germany, France, and Spain. Since 1996, the European Commission has funded different projects aimed at implementing and improving CLIL practices (Frigols Martin, Marsh & Naysmith 2007; Navés 2009). This is due to the fact that CLIL offered the perfect way of providing plenty of practice in the foreign language without increasing the number of language classes (Dale & Tanner 2012). From this, it followed that political and educational institutions at all levels started to take measures in order

to “unify and regulate the requirements which are to both support and guide CLIL professional development” (Frigols Martin, Marsh & Naysmith 2007: 35). In 2004, 80% of the European countries members of the Union were providing some sort of CLIL education (Dale & Tanner 2012). The *Action Plan* is one of the most important European developments regarding CLIL, as it established that all European citizens should be able to command, apart from their own mother tongue, two other languages in order to become multilingual citizens (Pérez-Vidal 2009). It is determined by this *Action Plan* that a foreign language be learnt as early as possible (nursery school); that this learning continue in primary education in an intensive and transdisciplinary way; and that a second foreign language be learnt at secondary school with an intensive and transdisciplinary character as well. Based on this idea, a number of initiatives have taken place in the last 20 years, as it is the case of the Spanish Ministry of Education/British Council Project in Spain (Llinares & Dafouz 2010).

However, in spite of the efforts made at the political and educational levels, as Dalton-Puffer (2011) claims, specific learning goals are missing in CLIL, due to the lack of investment of individual European countries in teacher training, CLIL implementation, and research. One of the exceptions to this situation, especially regarding research and implementation of programmes, is Spain. As far as research is concerned, a good number of studies have been carried out in the Basque Country (e.g. García-Mayo & García Lecumberri 2003; Ruiz de Zarobe 2008; Gallardo del Puerto et al. 2009); Catalonia (e.g. Escobar 2007; Escobar & Pérez-Vidal 2004); Madrid (e.g. Llinares & Whittaker 2010; Whittaker & Llinares 2009), and La Rioja (e.g. Jiménez Catalán et al. 2006). All these studies deal with different aspects of CLIL, such as the comparison between CLIL and non-CLIL students’ acquisition of different language aspects; CLIL implementation; CLIL methodology; and more recently, classroom-based research on students’ language use and development (Llinares & Morton 2010).

As a result of a commitment with the European policies that foster multilingualism, the implementation of CLIL programmes in Spain has grown rapidly in the last decade (Lasagabaster & Ruiz de Zarobe 2010). In spite of the fact that there is a national law of education that establishes the common framework, CLIL in Spain is heterogeneous and diverse, since there are 17 autonomous communities and they have the power to administer the educational system within each region (Lasagabaster & Ruiz de Zarobe 2010). There are two main contexts: monolingual communities, where Spanish is the official language, and bilingual communities (Catalonia, the Basque Country, Galicia, etc.), where Spanish coexists with another co-official language. These bilingual communities have been implementing bilingual education for more than 25 years and so they have offered a very good example “for the design and implementation of (CLIL) programmes in monolingual communities” (Lasagabaster & Ruiz

de Zarobe 2010: x). CLIL in Spain has moved, then, from regional to foreign language as the natural way "to generalise the use of more than one language as the medium of instruction" (Lasagabaster & Ruiz de Zarobe 2010: x). The CLIL approach in Spain is used in a wide range of subjects, depending on the programme, but one common denominator is that English is, overall, the most frequent language of instruction (Lasagabaster & Ruiz de Zarobe 2010).

In the region of Madrid, there are two existing projects at primary and secondary school levels: the MEC/British Council Project since 1996, and the CAM Bilingual Project since 2004 (Llinares & Dafouz 2010). Both institutions have made great efforts to implement CLIL, and many positive results have been reported (Llinares & Dafouz 2010). According to Llinares and Dafouz (2010: 100), one of the challenges in secondary education concerns teacher profiles: they are usually content teachers with little or no training in EFL, and so their awareness of the "specific language demands and characteristics of the different disciplines" needs to be raised. CLIL in tertiary education is more heterogeneous, but universities are rapidly developing English as a Medium of Instruction (EMI) programmes (Smit & Dafouz 2012; Dafouz & Smit 2017). All in all, what makes CLIL in Madrid different is its large dimension (over 500 state schools) and its fast implementation rate. As Llinares and Dafouz (2010: 110) point out, this fast growth of CLIL in Madrid makes research on the field very necessary.

Nonetheless, and in spite of the wide CLIL implementation in Spain and Madrid, there is a lack of CLIL pre-service programmes, and the in-service programmes which exist are insufficient (Fernández Fontecha 2009: 15). To this end (improving and/or designing pre- and in-service programmes), the present study could be a valuable contribution.

2.3 CLIL AND OTHER BILINGUAL EDUCATION MODELS: DIFFERENCES AND SIMILARITIES

This section will point out similarities and differences between CLIL and Canadian Immersion programmes and CBI in the US. The most important and straightforward similarity is that proficiency in an additional language is developed by teaching content through that target language (Llinares & Lyster 2014). In a similar way, the integration of content and language is a central aspect in the case of both immersion programmes and CLIL. Another similarity is that CLIL, CBI and FI could be called the "two for one approach", as they entail an increase of exposure to the target language that would be otherwise difficult to achieve, since school curriculum and timetables are at full capacity (Lightbown & Spada 2006). This increase of exposure has as a main source of support in Krashen's *Input Hypothesis* (1985), which

highlights the importance of the quantity of input (the more time a student is exposed to the foreign language, the better their learning will be) as well as its quality (it must be comprehensible and authentic, real input). Another commonality among the three programmes is that students have many opportunities for authentic and purposeful communication through the study of subject matter (Lyster 2007; Savignon 2004; Dale & Tanner 2012). In addition, the three of them give learners the opportunity of recognizing the value of culture (one of Coyle's four Cs), at the same time as they can develop cultural awareness and interest and respect for diversity (Savignon 2004). Finally, the three programmes are flexible, meaning that they vary greatly across contexts. They can vary in the entry point, the target language(s), or the content subjects taught through the target language (Llinares & Lyster 2014).

Although CLIL has clearly had the advantage of benefiting from previous and important research carried out in Canada and the United States (FI and CBI), differences must also be taken into account in the design of research studies and methodological proposals (Pérez-Vidal 2009). One of the main differences between CLIL, CBI in the US, and FI in Canada is that, in the case of CLIL, the target language is a foreign language, whereas the target language in CBI in the US is a second language, and in the case of Canada, it is another national language. The most important differences between CLIL and immersion programmes have been summarized by several authors as follows (Lasagabaster & Sierra 2009: 5-6; Dalton-Puffer 2011; Dale & Tanner 2012):

1. Immersion programmes are "carried out in languages present in students' context", whereas the languages of CLIL programmes are foreign languages and "many of the students only have contact with them in formal instruction contexts" (Lasagabaster & Sierra 2009: 5). CLIL classes demand interaction on the students' part with teachers and other classmates, but this interaction needs to be extended beyond the classroom for learning to consist in collaborative meaning-making (Donato 1994, Byram 1989). However, this practice outside the classroom is possible in CBI or in immersion programmes, but not in CLIL. Normally, students go out of the class and they do not find opportunities to use the foreign language.
2. In immersion programmes, most of the teachers are native speakers of French, whereas teachers in CLIL programmes are normally foreign language speakers of the target language.
3. Lasagabaster and Sierra also argue that students who enrol in immersion programmes normally start at a very early age, as opposed to CLIL programmes, in which sometimes students start in secondary education.

4. The teaching materials used in immersion programmes are aimed at native speakers, whereas CLIL teachers often use abridged materials.
5. The language objective in an immersion programme is to reach a near native proficiency, while CLIL programmes "cannot have such a far-reaching objective" (Lasagabaster & Sierra 2009: 6).
6. Immersion programmes are being held in a variety of languages, but the foreign language in CLIL is, in its vast majority, English. Less than or up to 50% of the curriculum is taught in the foreign language in CLIL programmes: one or more subjects are taught in the foreign language whereas the rest are taught in the L1. On the contrary, total immersion programmes work with the L2 all the time, except in the language arts mother tongue subject, and partial immersion programmes offer at least 50% of the curriculum through the target language.
7. CLIL lessons are normally timetabled as content lessons (Science, History, Geography, etc.). However, the target language continues as a subject taught by language specialists (traditional foreign language lessons).

Some of the ideas underlying these differences are defied by other authors. For instance, as for the first difference, Llinares and Lyster (2014) argue that, in spite of French being the other national language of Canada, English speaking families have little or no exposure to French. Therefore, the situation would be more similar than different from the CLIL case: in both contexts, students would not have much contact with the target language outside the classroom environment. In a similar way, Lasagabaster and Sierra's claim that students from CLIL programmes start at a late age (sometimes even in high school) is refuted by Llinares et al. (2012), arguing that in many Spanish CLIL programmes, such as those in Madrid or Andalucía, students start at a very early age (namely, the beginning of primary education). Contradicting Lasagabaster and Sierra, regarding language learning goals, Cenoz et al. (2014) argue that a native-like proficiency is not always the goal in immersion programmes, in the same way that there are CLIL programmes in which language learning goals are very high or even at a bilingual level. In any case, native-speaker competence does not seem to be the goal in neither of the programmes, but rather functional competence (Llinares & Lyster 2014): for instance, English speaking students in Quebec need French to function socially, and non-English speaking learners in Europe need English to participate in the European community, to travel, to do business, to study abroad, etc. Cenoz et al. (2014) also challenge the idea that CLIL and immersion programmes differ in the target language. Sometimes the target language in Immersion programmes is a second language (in the USA) and, in the case of CLIL programmes, the Eurydice report (2006: 8) states that the target language in CLIL is a "second

language (a foreign, regional or minority language and/or another official state language". Besides, regarding the limited contact that students can have with the foreign language outside the CLIL classroom, the country in which CLIL is practiced is a variable to be taken into consideration, as the amount of English used in Sweden, for instance, is more extensive than the use of French in some parts of Canada (Cenoz et al. 2014). Finally, Cenoz et al. (2014) also disagree with Lasagabaster and Sierra in that materials in immersion programmes are for native speakers: this is not always the case, as immersion teachers also adapt native materials, especially in early grades or in particularly difficult subjects. All in all, researchers seem to agree on three main aspects of CLIL: its European specificity, the vital importance of content and language integration, and the fact that the teaching and learning is carried out through a foreign language, as opposed to a second language (e.g. Dalton-Puffer et al. 2014).

2.4 CLIL METHODOLOGY: BENEFITS, CHALLENGES AND CONTENT AND LANGUAGE INTEGRATION

There are a great number of arguments in favour of CLIL and its methodology. However, the other side of the coin is that there are also a number of challenges in CLIL that need to be recognized and addressed in order to improve CLIL practices. In this section, special attention will be paid to the challenge of integration of content and language, as it is currently one of the main concerns among CLIL researchers (Llinares et al. 2012; Llinares 2015; Nikula et al. 2016).

2.4.1 CLIL methodology and its benefits

The advantages of CLIL, based on research findings, are summarized below:

1. Advantages for language learning:
    a. *Input-output*: CLIL increases the amount of exposure to the TL. Likewise, CLIL learners do not only receive a lot of input but also they have the opportunity of producing a lot of output (Pérez-Vidal 2009; Navés 2009).
    b. *Communication*: CLIL provides a naturalistic setting for language learning (as content-based programmes, Lyster 2007). It provides a purpose for the use of the second language in the classroom. Therefore, communication is meaningful, purposeful and authentic, language becomes then a tool for communication and not an end in itself (Dale & Tanner 2012; Pérez-Vidal 2009; Navés 2009; Llinares et al. 2012).

2. Advantages for integrated learning of content and language:
    a. *Learning and understanding*: CLIL learners' understanding of concepts is deeper, and their thinking skills and creativity are empowered. One of the resources students can use to understand concepts is linking new information in a foreign language with already existing knowledge in their mother tongue, which helps them transfer personal meanings from the L1 to the L2. In the same line, CLIL gives students the opportunity to engage in cognitively demanding tasks, which is another necessary condition for learning to occur (Coyle et al. 2010; Dale & Tanner 2012; Pérez-Vidal 2009; Dalton-Puffer 2009).
    b. *Language and content integration*: CLIL students learn the specific language and genres of the discipline (Llinares et al. 2012; Llinares 2015).
    c. *Meaning-oriented*: CLIL programmes emphasize meaning over form (Dalton-Puffer 2007, 2009; Llinares et al. 2012).
    d. *Communication*: in CLIL classrooms, meaningful interaction takes place at two levels, the level of the meaning necessary for language acquisition and the level of the language necessary for subject content learning (Llinares et al. 2012; Dalton-Puffer 2009). As Dalton-Puffer (2009: 197) puts it, CLIL "promotes learners' ability to communicate (effectively) in ways that traditional foreign language teaching does not", and it might also be added that this communication takes place in different contexts (formal/informal, general/specialized…). In relation to this, CLIL allows students to develop both BICS (Basic Interpersonal Communication Skills) and CALP (Cognitive Academic Language Proficiency) (Cummins 1981).
3. Advantages for learning in general:
    a. *Learners' motivation*: CLIL motivates learners, fostering intrinsic motivation to use the foreign language in order to communicate, which makes unconscious learning possible (Pérez-Vidal 2009).
    b. *Learners' internationalization*: CLIL learners are more prepared for going abroad and study in another language (Navés 2009; Pérez-Vidal 2009).

Most of the advantages mentioned above are shared by other types of bilingual education. For instance, the advantages for language learning and for learning in general can be also found in FI or CBI, as well as the characteristic of meaning being emphasized over form (Cenoz et al. 2014). In addition, some features such as the advantage of learners being intrinsically motivated can, of course, be found in other types of methodologies, not necessarily related to bilingual education models. In turn, the advantages described in the second group can be said to be CLIL-specific, not necessarily meaning that they actually happen in all "so-called" CLIL programmes,

but the integrated teaching of academic content and foreign language is what characterizes CLIL, at least in principle (Pérez-Vidal 2009; Dalton-Puffer et al. 2014).

The advantages of CLIL programmes described mainly in the first group make the conditions for FL learning more similar to those found in L1 acquisition. Those are the conditions that have been found the most adequate for second language learning, in particular, that meaning is stressed over form, that input is comprehensible (at the students' level or a bit beyond) and that students have opportunities to engage in meaningful interaction in a context free of anxiety (Krashen 1982; Lightbown & Spada 2006; Long 1990; Swain 2000). Indeed, CLIL learners supposedly learn the foreign language more quickly and achieve higher levels of proficiency than their mainstream peers, also with less chance of fossilization (Dale & Tanner 2012). In spite of the emphasis placed on meaning over form, Lyster (2007) advocates for a counter-balanced approach in immersion programmes, a stance that has also been adopted, from different perspectives, by other researchers in CLIL (Dalton-Puffer 2009; Whittaker & Llinares 2009). As already mentioned, the methodology used in this study will also account for language-focus and content-focus interaction in primary CLIL classrooms.

For CLIL language teaching to be successful, CLIL programmes need to have a series of characteristics (some of which are the object of analysis of this investigation). This includes actively engaging students through communication, offering scaffolding, monitoring students' progress, providing immediate feedback, and enhancing collaborative and autonomous learning (Navés 2002; 2009: 34; de Graaff et al. 2007). These are also relevant characteristics of AfL or Formative Assessment (namely meaningful, authentic and scaffolded interaction, immediate feedback, and collaborative and autonomous learning), as we will see and explain further on in the next chapter.

CLIL provides benefits which could be difficult to find in monolingual education, for students, teachers, and schools. Regarding students, some of the benefits are (Wolff 2007: 19-22):

- The CLIL learner forms original concepts which have not been formed previously in their mother tongue. According to Wolff (2007: 19), "concept formation […] is independent of the learner's mother tongue".
- CLIL learners learn the language in a better way, since they have more input in the foreign language, at the same time as they process it in a deeper way (Coyle 2007; Lorenzo et al. 2009; Ruiz de Zarobe & Jiménez Catalán 2009). In CLIL, content serves as a "scaffold for the language learning process" (Wolff 2007: 20).
- CLIL learners also learn the content better, since they are required to make harder cognitive efforts when the content is presented in a foreign language. Yet, this last

affirmation needs to be proven true by more studies, as research in this area does not seem to be entirely conclusive (Seikkula-Leino 2007; Badertscher & Bieri 2009; Admiraal et al. 2006; Airey 2009; Lim Falk 2008; Van de Craen, Ceuleers & Mondt 2007; Anghel, Cabrales & Carro 2016).

- They are more prepared for working life, since they learn the genres and registers of academic subjects (formal registers) and they can communicate effectively in those registers.
- They are prepared for the future demands of their profession, since they practise strategies such as working in groups, solving problems autonomously, presentation techniques.

As far as teachers and schools are concerned, implementing CLIL can be a trigger for innovation, renewal, and reflection in a school (Dale & Tanner 2012: 14). Also, non-native teachers teaching in CLIL programmes (which is the usual case in most contexts such as Spain) are able to improve their L2 proficiency and be aware of the linguistic aspects of their subject, as well as be exposed to new ideas about their subject and how it can be best learnt and taught (Dale & Tanner 2012). CLIL teachers benefit from CLIL methodology in that they are able to develop and improve a lot of skills they may or may not already have. Among the skills CLIL teachers need to have, they need to know how language is used in their subjects so that they can help learners notice this use and overcome linguistic problems and challenges (Dale & Tanner 2012; see also Llinares, Morton & Whittaker 2012). In addition, the CLIL teacher needs to activate students' previous knowledge about the topic at hand and provide appropriate and multimodal input. Another skill they need to have is to encourage pupils to interact, for which they will have to use good questioning techniques. Dale and Tanner (2012) also claim that CLIL teachers need to assess students' progress in relation to both content and language, giving them useful feedback that will help them develop in both areas. It could be added that CLIL teachers need to engage students cognitively: there has to be a cognitive challenge, but also appropriate support or scaffolding, which will decrease as students make progress (Coyle 2006). Again, to do that, appropriate questioning and feedback needs to be offered by teachers. In the same way, teachers have to raise students' meta-cognitive processes: they have to be aware of how they learn to learn, and teachers can encourage students' awareness of their learning processes through the use of certain types of questions. The relevance of questions and feedback in CLIL classroom interaction will be brought to the fore in the present study.

2.4.2 CLIL methodology and its challenges

As well as benefits, there are also different challenges that need to be addressed in CLIL: methodological challenges, teacher challenges. and student challenges. There are two main methodological challenges: a) how to move away from transmissional models of teaching, and b) the issue of students' cognitive level mismatching their level of knowledge of the foreign language.

Moving away from transmissional models of teaching (Barnes 1975; Alexander 2004; Lyster 2007) is especially challenging in some countries like Spain, where different subjects that are now CLIL content subjects have been traditionally taught following a transmissional model (Llinares & Pascual 2014). If CLIL is to be successful, it should give way to more inquiry-oriented models of teaching and learning, since the development of thinking skills and the realization of cognitively demanding tasks are two important characteristics of this educational model. CLIL classrooms demand a high level of interaction if its characteristics and advantages are to be fulfilled (see section 2.4.1 above for its characteristics and the advantages for language learning, for learning in general, and for integrated learning of content and language) (Coyle 2007; Dalton-Puffer 2009). Therefore, CLIL lessons need to move from teacher-centred practices to more student-oriented ones (Nikula et al. 2013). In this vein, Llinares and Whittaker (2009) advocate for CLIL as a context in which L2 literacies are developed and students participate in CLIL classrooms through using the language to carry out certain academic language functions through which subject meanings are expressed. Therefore, CLIL students need to be given space and opportunities in interaction to articulate their understandings. The importance of CLIL as a space for interaction and inquiry has been theorized in the literature and also proven at the practical level in different research studies (Llinares & Whittaker 2009; Llinares et al. 2012; Nikula 2005; Moore 2011; Dalton-Puffer 2007; Llinares & Pascual 2014).

The second challenge is related to the mismatch between students' language and cognitive level. CLIL involves a double aspect: learning to use a foreign language appropriately, and using that language to learn concepts effectively. This leads us to another methodological challenge: the possible mismatch between students' cognitive level and their level of language knowledge (Coyle, Hood & Marsh 2010). In many cases, their cognitive level is superior to their language level. According to Coyle, Hood and Marsh (2010: 43; based on Cummins 1984), when the cognitive level is too low, learning is limited, whereas if the language level is too high, effective learning cannot occur either. Therefore, for effective learning to take place, each individual student needs to be cognitively engaged at their own appropriate level. The following chart

(Cummins 1984) shows the different possibilities that we can find when combining cognitive and linguistic demands:

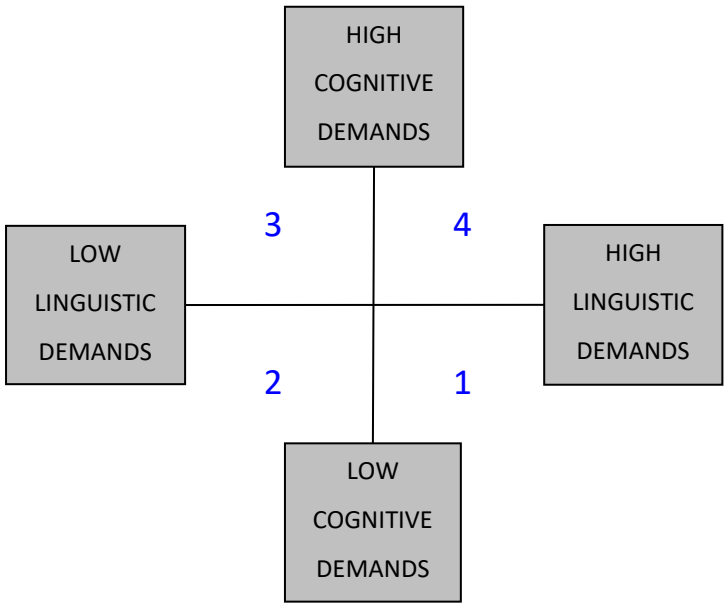


Figure 2.1 Cognitive and linguistic demands (Coyle et al. 2010: 43; adapted from Cummins 1984).

When the learner does not have a high level of the language, teachers would have to focus on quadrant 3 to make sure that learning still takes place. When students make progress in their language level, then the teacher can move to quadrant 4. Quadrant 2 limits development, since it does not require challenging at the linguistic or cognitive level. Finally, quadrant 1 can be used when focus on form is crucial to continue progressing in learning, because it requires high linguistic demands but low cognitive ones. One way of engaging students both linguistically and cognitively is through the use of a variety of questions, as this study will show. As a complement to the chart above, Cummins (1992) proposed a paradigm in which language tasks can be context-reduced or context-embedded on the one hand, and cognitively demanding or cognitively undemanding, on the other, creating in this way another quadrant diagram.

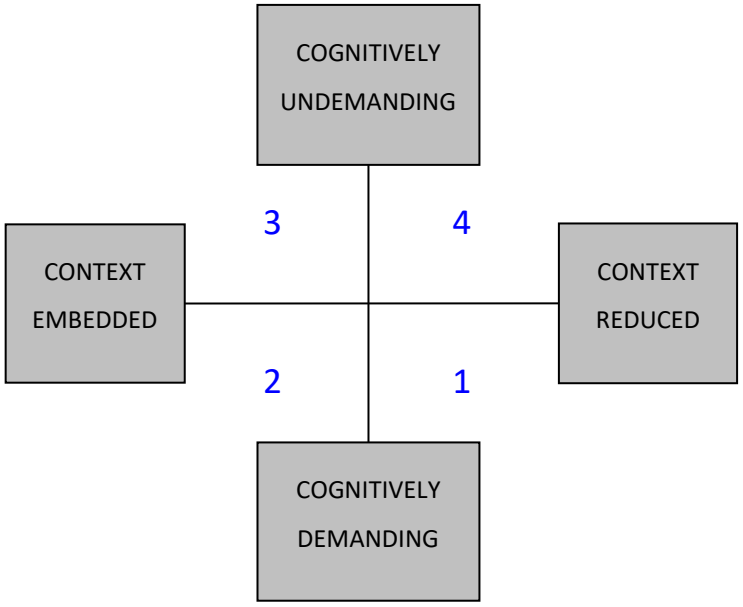


Figure 2.2 Cummins' diagram of language tasks (1992).

When teachers provide many contextual cues, they may be turning the task into cognitively undemanding. When content and language are integrated, language can be practiced in more cognitively demanding tasks (Snow et al. 1992). Comprehensible language is rich in contextual cues (objects, visual aids, gestures…). Less comprehensible language which is characterized by having fewer contextual cues and comprehension usually relies on the listener/reader's ability to infer meaning from text without the help of non-verbal cues. Comprehension is also affected by the second dimension of the diagram: task complexity. Tasks involving language use can be categorized into: easy and contextualized, difficult but contextualized, context-reduced but easy (leading to rote learning), context-reduced and difficult (Cummins 1992). The quadrants above can help teachers identify mismatches between students' cognitive and linguistic levels and propose the appropriate tasks to keep them engaged and to promote further learning.

Moving on to the challenges regarding CLIL teachers, one of the most important ones is that CLIL will require different teachers from different areas to work cooperatively (Llinares et al. 2012: 21; Dale & Tanner 2012). In de Bot's words (in Marsh 2002:32),

> it is obvious that teaching a subject in a foreign language is not the same as an integration of language and content… language teachers and subject teachers need to work together… [to] formulate the new didactics needed for a real integration of form and function in language teaching.

One potential uniqueness of CLIL, that of content and language integration, can also be one of its potential weaknesses. It is believed that when the teacher is a specialist in language, some

contents could be ignored or weakened; on the other hand, if the teachers are specialists in content, maybe the linguistic demands could be in danger (Coyle, Hood & Marsh 2010). Teachers, whether language or content specialists, find themselves in a challenging situation: content teachers because they are not familiar with the language demands of their subject; language teachers because they may be less familiar with scientific, subject-specific concepts (Dale & Tanner 2012). What seems clear is that teachers (no matter their speciality) must be able to identify the language-specific needs of their subject. For that, mere guidelines and lists of vocabulary are not sufficient. They have to be aware of key language, grammatical features, and also the genre required by the subject (Llinares et al. 2012). As Dalton-Puffer (2007:295) argues:

> At present, at least in Austria, a CLIL curriculum is defined entirely through the curricula of the content subjects, with the tacit assumption that there will be incidental language gains. But why should we be doing CLIL at all if there are no language goals present*?*

A third challenge relates to CLIL students. These learners face the challenge of having to learn content through the use of a foreign language, which involves both cultural and linguistic demands at various levels (word, sentence, discourse levels) (Dale & Tanner 2012). Along with this, students may also have to face affective challenges, since speaking in a foreign language may create situations of anxiety when using it in front of the teacher and classmates, which could affect students' self-confidence.

2.4.3 Integrating content and language, the biggest challenge

As already pointed out above, one of the main challenges in CLIL is the integration of content and language. This integration is paramount, as many researchers have claimed (Dalton-Puffer 2011; Lyster 2007; Llinares et al. 2012), and there is growing concern regarding its conceptualization (Gajo 2007; Dalton-Puffer 2011; Llinares 2015; Nikula et al. 2016). Dalton-Puffer (2011) proposed to stop looking at content and language as two separate entities and instead see them as one process. The understanding of integration entails the application and combination of constructivist and sociocultural theories of learning, Systemic-Functional Linguistics, discursive pragmatics, and even SLA models such as task-based language learning (Llinares 2015). Some models that have been proposed for the study of content and language integration are Coyle's (2007) 4Cs and functions of language framework, Lyster's (2007) counterbalanced approach, and Llinares et al.'s (2012) and Llinares' (2015) application of SFL in combination with other models.

Working on CLIL and its specific features, Coyle (2007) designed a conceptual triad that integrated content and language learning. This triad consists of *language of learning*, *language for learning*, and *language through learning*. These three functions exemplify how a foreign language can be used for the construction of content knowledge in CLIL. *Language of learning* is the language students need to have access to the basic concepts and skills of the topic. *Language for learning* is the kind of language needed for communicating effectively in the foreign language. Finally, *language through learning* is when learners are encouraged to articulate their understanding. In so doing, and as they acquire new knowledge, they will also acquire new areas of meaning. New language emerges from or through learning (Coyle, Hood & Marsh 2010). Students' articulating their understandings (language through learning) is crucial for both language and content learning, which is one of the reasons why in CLIL classes there is a higher demand of talk and interaction than in traditional classrooms. In this line, Met (1998: 34) claims that:

> students need to communicate with the teacher, one another, or texts, in order to access or apply content. In so doing, the cognitive demand of task requires students to call upon their existing knowledge, concepts, skills and strategies. This strengthens the connections between the elements of language being practiced/ learned and previous knowledge. As we have seen, research indicates that strengthening and making connections amongst concepts and knowledge increases learning and retention.

A second model designed by Coyle (2007) for the exploration of content and language integration is the 4 C's framework. According to Coyle, Hood & Marsh (2010), there are four main factors playing a role in CLIL (the 4Cs): subject matter (*content*); language learning and use (*communication*), which goes beyond the grammatical rules and involves using language in various ways (learning to use language and using language to learn); learning and thinking processes, engaging students to construct their own understanding and be challenged (*cognition*); and, finally, *culture*, as language reflects culture and culture determines how we see and interpret the world. The inclusion of culture also contributes to making students be tolerant and aware of the "otherness" to understand the "self".

Lyster's counterbalanced approach, rooted in the necessity of taking up and systematising focus on form opportunities in content classes (see section 2.1.1), has been advocated for in CLIL contexts precisely because it could help integrate content and language (Llinares et al. 2012; Nikula et al. 2013; Morton 2012). To do this, language objectives should be present in the curriculum and developed hand in hand with content ones (Llinares et al. 2012). In order to improve this aspect, the concepts of *content-obligatory language* (necessary for understanding

subject matter) and *content-compatible language* (language that can be taught through content when the opportunity comes) could be helpful in order to identify clear language goals in the curriculum (Snow, Met & Genesee 1989).

Finally, Llinares et al. (2012) also dealt with integration in CLIL, drawing on Systemic Functional Linguistics, Vygotsky's sociocultural theory of learning, and social models on second language acquisition. The three of them share the idea that language and learning are social processes. Llinares et al. (2012), then, developed the following framework in order to describe the different roles of language in CLIL:

1. *Subject literacies*: that is, the different genres and registers through which content knowledge is expressed.
2. *Classroom interaction*: in which we find three elements: focus (the content being talked about and its purposes), which identifies instructional and regulative registers; approach (communication systems); and action (interaction patterns and scaffolding).
3. *Language development* of students' linguistic competences through classroom interaction and their engagement with the genres and registers through which content knowledge is realized. Language development includes three dimensions: expressing ideational meanings (lexico-grammatical features to express key ideas and concepts of the subject), expressing interpersonal meanings (those related to social relationships and attitudes) and expressing textual meanings (moving from more spoken-like forms of language to more written-like forms of language, typical of academic school subjects).

More recently, Llinares (2015) expanded the model incorporating discursive pragmatics and TBLT as compatible models with SFL for the understanding of content and language integration.

This section has presented the benefits that a CLIL methodology supposes for students, teachers, and schools. In addition, the main challenges in CLIL have also been addressed. If CLIL practice is to improve, challenges need to be tackled and turned into strengths.

2.5 CLIL RESEARCH

Research on CLIL has drawn on different theoretical models: SLA, SFL, sociolinguistic, sociocultural, discourse (see Llinares & Morton 2017 on different approaches to CLIL research). SLA applications to CLIL and bilingual education have been the focus of interest for decades. The idea that language learning can happen without formal instruction, in a naturalistic way, in CLIL classrooms is partly based on Krashen's (1985) Input Hypothesis and the

assumption that when learners are exposed to comprehensible input, acquisition will occur. Swain's *Output Hypothesis* (1995) has equally inspired CLIL research. This theory puts forward the idea that learner production of modified output ("comprehensible output") is necessary for L2 acquisition. Thirdly, through interaction and negotiation of meaning, learners are given the opportunity to modify their output, focus on their interlanguage grammar, notice the gap in their language system and create more complex target-like forms (Morris 2002). Focus on form has been a further development, where special attention is paid to formal aspects of language as carriers of meaning (Dalton-Puffer 2011). In immersion contexts, Lyster (2007) has proposed a counterbalanced approach, putting equal emphasis to focus on meaning and focus on form, moving away from the idea of implicit and naturalistic language learning (Krashen 1985). However, in Europe, there has been little research directed towards form-focused activities in CLIL lessons (but see, for example, García Mayo & Basterrechea 2017). Other researchers have dealt with a broader "focus on language", relating form to meanings and functions (e.g. Llinares et al. 2012).

In addition to SLA approaches, CLIL research has also made use of theories that view learning as a social construction, which depends heavily on context (Dalton-Puffer 2011). According to these theories, such as Systemic Functional Linguistics or Sociocultural Theory, humans learn a language by participating in the social world they live in and interacting with others. Following these theories, content-based situations take students' attention close to meanings conveyed through language, emphasising, in this way, the relations between meaning and form (Dalton-Puffer 2011).

2.5.1 Lines of research in CLIL

Research on CLIL has revealed that CLIL classes are similar to other types of lessons (such as FL or L1 content classes) in that it is educational discourse (Nikula, Dalton-Puffer & Llinares 2013). However, differences between FL and CLIL lie on the presence of content pedagogy: language learning opportunities which are difficult to orchestrate in FL classrooms arise from the fact that "participants clearly interpret their roles differently and thus orient differently to their respective tasks" (Nikula, Dalton-Puffer & Llinares 2013: 81).

One area that has received very little attention is the effect of CLIL on content learning. When CLIL started to be implemented, there was fear among educators and parents about whether (and how) content learning would be affected. Students' knowledge of the foreign language is, evidently, poorer than their knowledge of the L1. Thus, much concern is placed on reduced subject competence and simplification of content (Dalton-Puffer 2011). In general, this type of

investigation is less conclusive than CLIL research in other areas. There are studies indicating positive results (Vollmer et al. 2006; Van de Craen, Ceuleers & Mondt 2007), while others are more critical and point out that CLIL learning leads to less content learning (Lim Falk 2008; Airey 2009; Anghel, Cabrales & Carro 2016). Likewise, other studies report neither a positive nor a negative effect of CLIL on content learning (Admiraal et al. 2006; Badertscher & Bieri 2009; Jäppinen 2005).

Another area of CLIL research that is growing is the effect of the use of the L1. There is controversy in this issue as there are some advocates for the no L1 policy, whereas a growing number of researchers that L1 use can also be positive (Brooks, Donato & McGlone 1997; Moore 2002; Nikula & Moore 2016; Ohta 2001). These authors claim affective and strategic reasons for the use of the L1. Furthermore, they argue that this use is positive for the development of L2 skills. Mori (2002) states that hindrances in communication can take place if there is a target language only policy. Besides, she stresses the fact that age and foreign language level of the learners should be taken into account too.

Apart from the two research areas mentioned above, there are currently two relevant lines of research within CLIL: studies dealing with policy and implementation of CLIL programmes and methodologies (Mehisto, Marsh & Frigols 2008; Coyle et al. 2010), and empirical studies on students' L2 gains and development (Dalton-Puffer 2007; Ruiz de Zarobe & Jiménez Catalán 2009; Lorenzo, Casal & Moore 2009). In the first branch, that of policy and CLIL implementation, different CLIL projects have been funded by the European Commission, and efforts are being made in order to unify the requirements that guide and support CLIL development. However, since there is no official policy regarding CLIL implementation, different EU member states have launched diverse initiatives since the 1990s, both at national and local or regional levels (Frigols Martin, Marsh & Naysmith 2007; Lasagabaster & Ruiz de Zarobe 2010). Within examples of good practice, we can find different projects being implemented in countries such as Finland, Spain, Austria, Germany and France (Baetens Beardsmore 2007). Since there is not a unique CLIL model that suits all CLIL contexts, it is therefore necessary to create communities of practice that share principles, aims and values on CLIL methodology and to establish a dialogue and cooperation among different European countries in order to raise the quality of CLIL education (Coyle 2007). Communities of practice would also help raise the quality of teacher education in CLIL, which is one of the key aspects and main concerns regarding the success of this educational model (Coyle 2007; Frigols Martin, Marsh & Naysmith 2007; Lasagabaster & Ruiz de Zarobe 2010).

In the second CLIL research branch, empirical studies, we can differentiate several lines of investigation:

- Product-oriented studies on students' learning regarding linguistic aspects such as syntax, vocabulary, pronunciation (Ruiz de Zarobe & Jiménez Catalán 2009); and product-oriented studies on learner linguistic outcomes related to a series of factors, such as teacher characteristics, time devoted to bilingual education, subject taught in English (Lorenzo et al. 2009).
- Process-oriented studies focusing on CLIL classroom discourse, paying special attention to pragmatic aspects (Dalton-Puffer & Nikula 2006; Llinares, Morton & Whittaker 2012; Nikula et al. 2013).
- Analyses of students' written and spoken production in relation to subject disciplines (e.g. Llinares & Whittaker 2007; Llinares & Whittaker 2009).
- CLIL students' attitudinal factors (Lasagabaster & Sierra 2009).

Studies on foreign language learning have been prominent in CLIL research from the beginning, adopting a product perspective, and even some studies that analyse classroom discourse take a foreign and second language learning position (Nikula, Dalton-Puffer & Llinares 2013). Others, on the contrary, adopt a process-oriented stance of language learning, meaning that learning is thought to occur in interaction informed by pedagogical goals (Nikula, Dalton-Puffer & Llinares 2013). Bilingual education should not only be seen from the productivity perspective (students' attainment) but also from the perspective of how teachers and students use language in classroom activities, which is key for the understanding of language and curriculum content learning in specific local contexts and, at the same time, this can be fed back into pedagogy and policy issues (Leung 2005a: 238). The present study can be classified as process-oriented and situated within empirical studies at the micro dimension (Dalton-Puffer & Smit 2007), that is, it focuses immediately on the participants and the process of interaction and learning in which they engage.

More studies are needed that specifically address the challenge of content and language integration. In the past few years, such studies have grown, showing the special duality of CLIL and the growing concern among researchers about this issue (Llinares et al. 2012; Nikula et al. 2016). Systemic Functional Linguistics has turned out to be very helpful for such analyses. SFL perspectives have been used to carry out studies on the ideational, interpersonal and textual functions of language (Llinares, Morton & Whittaker 2012; Llinares & Morton 2012; Whittaker, Llinares & McCabe 2011; Morton & Llinares 2016). Regarding the ideational function, some of the findings reveal that CLIL students use the type of processes and

circumstances that are expected in the different genres of different subjects (History and Geography) (Llinares & Whittaker 2009). However, when compared to their peers in L1 subject lessons, L1 students were more proficient in the use of academic register than CLIL students (Llinares & Whittaker 2010). As for the textual function, CLIL students also have difficulties to manage referential elements in texts (Whittaker et al. 2011), and finally, in relation to the interpersonal function, students' use of appraisal highly depends on the type of task (Llinares & Dalton-Puffer 2015) as well as on students' competence level (Morton & Llinares 2016).

On the integration of language and content in CLIL, special mention needs to be made of Llinares, Morton and Whittaker's study (2012) on interaction and its role in the learning of both components in integration. They show how, in the instructional register, talk moves from everyday experiences to scientific knowledge, in the same way that students' foreign language use needs to move from everyday and context-embedded language to academic and written language. In this study, the authors also illustrate the opportunities that a dialogic-interactive mode of communication offers in order to engage CLIL students cognitively. In terms of IRF patterns, the study emphasizes and instantiates a variety of scaffolding strategies that CLIL teachers can use to support students' development of linguistic resources in the foreign language in order to express different kinds of meaning relevant for the subject.

All in all, there is great heterogeneity in CLIL research, making conclusions difficult to make (Nikula, Dalton-Puffer & Llinares 2013). Both quantitative and qualitative approaches, as well as process-, product- and participant-oriented research are called upon (Nikula, Dalton-Puffer & Llinares 2013). In CLIL research, it is important to combine different approaches in order to comprehend the multifaceted nature of the CLIL classroom, as has been advocated by some researchers (Nikula, Dalton-Puffer & Llinares 2013; Llinares, Morton & Whittaker 2012).

2.5.2 Research comparing EFL and CLIL: different opportunities for language learning

All types of learning are social and dialogic in nature, and CLIL classes are just one more variant of the different types of classrooms that we find in educational contexts (Dalton-Puffer & Nikula 2006; Dalton-Puffer & Smit 2007; Dalton-Puffer 2009). If we think of education in general, every class type shares some features, namely they are arranged in lessons taking place in schools (or universities), hence sharing characteristics such as participant roles, goals, time structure, etc. (Dalton-Puffer 2011). These common characteristics that all classroom types share make CLIL students already familiar with the school idiosyncrasy and with the cultural context. Within this common educational framework, however, CLIL has its own specificities:

> My claim, then, is that CLIL provides a space for language learners that is not geared specifically and exclusively to foreign language learning but at the same time is predefined and prestructured in significant ways by being instructional and taking place within the L1 matrix culture. This, I claim, is a significant source for the self-confident and self-evident use of the foreign language and its ultimate appropriation by many CLIL learners, which is regularly observed to be the most striking outcome of CLIL programs (Dalton-Puffer 2011: 195-196).

The comparison between bilingual programmes and regular second language lessons has rendered a lot of research. In fact, most of the research in CLIL focusing on attainment in the target language has been based on the comparison between CLIL and EFL students. It is to be expected that CLIL students outperform their peers studying the target language in traditional EFL classrooms, mainly because CLIL students have both CLIL classes and EFL ones (Dalton-Puffer 2011). This expectation is confirmed by several surveys (Lasagabaster 2008; Lorenzo et al. 2005; Ruiz de Zarobe 2008). Different studies show specific areas in which CLIL students outperform students studying in regular programmes:

- Within *vocabulary*, it can be said that CLIL students' receptive and productive lexicon is larger, has a wider stylistic range, and is used more appropriately (Jexenflicker & Dalton-Puffer 2010; Ruiz de Zarobe 2010; Jiménez Catalán & Ruiz de Zarobe 2009).
- Regarding *writing*, some studies have shown that CLIL students have more morphosyntactic and lexical resources at their disposal (Ruiz de Zarobe 2010; Jexenflicker & Dalton-Puffer 2010). Other dimensions of writing where CLIL pupils outperformed their peers were accuracy and pragmatic awareness (Dalton-Puffer 2011). The writing aspects where CLIL teaching seems to have no effect on are those beyond the sentence level (cohesion, coherence, genre, register) (Dalton-Puffer 2011).
- An area where CLIL instruction seems to have no positive impact is *pronunciation* (Gallardo del Puerto, Gómez Lacabex & García Lecumberri 2009).
- The main area where a most noticeable difference between CLIL and mainstream learners has been found is in their *spontaneous oral production*. Many studies corroborate this finding, mainly that CLIL students have greater fluency and self-confidence when speaking spontaneously (Maillat 2010; Moore 2009; Dalton-Puffer et al. 2008; Mewald 2007; Nikula 2008).
- With respect to *teacher-student interaction*, CLIL research has mainly focused on code switching (Dalton-Puffer 2007; Nikula 2007a), IRF patterns (Dalton-Puffer 2007; Nikula 2007b), collaborative interaction (Moore 2011), and repair (Dalton-Puffer 2007; Serra 2007; Smit 2007). Repair in CLIL classrooms, for instance, seems to be lower

than in EFL, and that may be why students talk more freely (Dalton-Puffer 2009). There is a preference for self-repair (as in non-educational talk), but other-repair and other-initiated repair are also present in CLIL classrooms (Dalton-Puffer 2009). Generally speaking, there is consensus that CLIL learners demonstrate good interactive skills and that CLIL experience has a positive impact on students' oral capacities (Moore 2011). However, Moore (2011) calls for the need of homogeneity and using more robust and less fuzzy concepts when measuring oral output.

- With regard to *sociolinguistic competence*, CLIL classrooms seem to have no advantages over EFL classes (Dalton-Puffer 2009).
- Echoing studies on *negotiation of meaning* in immersion classrooms, studies on negotiation of meaning comparing CLIL classrooms with EFL classrooms conclude that the number of negotiation of meaning sequences are double in CLIL classrooms (Badertscher & Bieri 2009; Lochtmann 2007).

In sum, research carried out so far on CLIL when compared to EFL classes shows the superiority of CLIL students over their mainstream peers in almost all the areas investigated. Specifically, CLIL students seem to be particularly strong in strategic competence (Canale & Swain 1980), as they are able to convey content at early stages when their linguistic resources are still limited (Dalton-Puffer 2011). However, more comparative studies between CLIL and L1 content classes are still needed in order to report more conclusive evidence regarding CLIL students' content learning (but see, for example, Llinares & Whittaker 2010).

2.6 THE LANGUAGE OF THE DISCIPLINES IN CLIL

The different linguistic features, registers, and genres of some CLIL subjects have been already described in the literature (Dale & Tanner 2012: 48-83). Following Dale and Tanner (2012), the following list describes the subjects that will be object of analysis in the current study: Arts, Geography, History, Science, and Drama. Citizenship is the only subject that appears in this investigation that has not been researched at all and therefore its linguistic features, registers and genres cannot be pre-specified.

1. *Arts*: language is used to describe, explain, and evaluate objects and techniques. Visual experiences are as important as cognitive ones. Learners can learn creative and abstract language, as well as the language of emotion and expression. There are many hands-on activities and the input tends to be informal. The fact that most work in art is done through materials rather than language means that teachers have to provide

opportunities for language work (questioning, hypothesising, evaluating, writing, social talk, language for thinking about complex issues, etc.).

2. *Geography*: language is used to explain, study, and analyse the earth. Its input is usually multimodal. Students need to speak, write, and think like geographers and, to do that, they need to understand the information and the specific vocabulary, organize the information, and use language for thinking skills. Geography offers good opportunities for language work. Geography genres are classified into: geography reports, mainly descriptive, which introduce learners into geographical phenomena, with specific terms and definitions; and explanations, reflecting relations between phenomena (sequential, causal, consequential) (Llinares et al. 2012).

3. *History*: language is used to narrate, explain, and analyse past events. In this subject, pupils will have think, talk and write as historians. Input also tends to be multimodal, relying a lot on written texts. History is also a good discipline through which linguistic aspects can be worked: students not only have to narrate, but also evaluate sources, argue viewpoints, and look at events from different perspectives. As for History genres, there is a first classification between non-chronological genres, recording genres, explaining genres, and arguing genres, from least to most difficult. Within non-chronological genres, period study is the most frequent one, and it describes the characteristics of a historical period. Within recording genres, which are organized through time, there are historical recounts, in which events are told in a sequence following a chronological order; and historical accounts, already include explanations and causes why a sequence of events happened. Within explaining genres, we find historical explanations, which attend to causes and consequences of events. Finally, historical argument within arguing genres is the discussion of interpretations of historical events or figures (Christie & Derewianka 2008; Coffin 2006; Schleppegrell 2004; Llinares et al. 2012).

4. *Science*: language is used to describe, explain, and analyse scientific phenomena. Input is also multimodal in order to support different ways of learning. In Science, students develop language for thinking skills, such as reasoning, hypothesising, questioning, problem-solving… Learners learn to speak, think, and write like scientists and, as the content becomes more complex through years, learners need to become proficient at expressing complex and abstract ideas in a formal and academic way both when speaking and when writing. The different genres in Science are: procedures, which are concerned with instructing students how to carry out an experiment and its

correspondent procedural recounts, in which students write up experiments; science reports (descriptive or classifying), which organize scientific knowledge; and science explanations (causal, consequential, factorial), which provide explanations for different scientific processes and phenomena (Llinares et al. 2012).

5. *Drama*: language is used to understand, create, evaluate, and perform their own and each other's work. Therefore, there are a lot of hands-on activities. It is a good subject for introducing students into CLIL, because the new language introduced is always supported with visual and physical support. It gives students the possibility of using language in a varied way, and they can experiment with the expression of emotions. Drama also lends itself to reading and writing different texts as a way of fostering linguistic creativity.

The above described subjects can be classified into two groups: more practical and experiential subjects, like Drama and Arts, and more scientific and analytical ones, like Science, History and Geography. As already pointed out, Citizenship has not been researched yet and so it cannot be pre-classified. In this dissertation, the subject Science includes both Natural and Social Sciences, that is, Science, History, and Geography.

2.7 SUMMARY AND CONCLUSION

This chapter has been devoted to CLIL as an educational practice and a form of bilingual education. In order to better understand this educational model, it has been related to other non-European but CLIL-related models (Canadian Immersion programmes and CBI programmes in the US). The situation of CLIL in Europe, Spain, and Madrid has been described before moving into CLIL methodology in detail: its benefits, its challenges, and the integration of content and language. Next, CLIL research has been thoroughly explored: existing theoretical models, the different lines of research existing in CLIL, and the comparison between CLIL and EFL classes. Finally, the last section has described the role of language in the different disciplines that are present in this study. It seems that research carried out up to now has proved that CLIL has positive effects in students' command of the foreign language, as well as in their cognitive capacities. Nonetheless, still further research is needed in many areas in order to better understand and implement this educational practice. One of these areas is assessment in interaction/interaction in assessment, to which this dissertation will hopefully contribute.

3

Theoretical framework

This chapter presents the main theoretical perspectives needed to inform the present study, namely two: classroom interaction and language learning; and assessment for learning (AfL). Each section is formed of subsections in order to discuss different theoretical aspects relevant for the present study. The third section of the chapter will address the role of classroom interaction in AfL, which combines the two models applied in this study.

3.1 CLASSROOM INTERACTION[1] AND LANGUAGE LEARNING

The study of classroom discourse is the study of the language that teachers and students use to communicate with each other in the classroom (Zhang Waring 2008). It is through spoken discourse that teaching and learning take place and that participants' identities are constructed. Classroom interaction has been thoroughly studied and different studies have approached the topic from different perspectives (see Sinclair & Coulthard 1975; Walsh 2006; Christie 2002; Mercer 1995; Edwards & Westgate 1994). Some researchers have focused their analyses on triadic dialogue (IRF patterns, see section 1.1.2 below), paying attention to larger stretches of discourse such as exchanges, sequences, or episodes (see Sinclair & Coulthard 1975; Wells 1993; Nassaji & Wells 2000; Lemke 1990; Cazden 1986, 2001). Others, coming from the Conversation Analysis (CA) tradition, have provided detailed descriptions of classroom interaction as a space in which intersubjectivity is developed and participants' intentions are crucial (see Walsh 2006, 2012). Classroom discourse has also been studied from the Systemic Functional Linguistics perspective, in which meaning-making is affected by the speakers' linguistic choices (see Christie 2002). Analyses of classroom interaction coming from sociocultural perspectives have stressed the key role of mediation and scaffolding (see Mercer

[1] Classroom discourse and classroom interaction are used as synonyms in this dissertation.

1995; Edwards & Westgate 1994; Aljaafreh & Lantolf 1994). Finally, second language acquisition studies on interaction have focused on negotiation of meaning and how interaction helps second language learning and development (see Long 1996; Lyster & Ranta 1997; Ellis 1997; Sheen 2004). There are two main research traditions in the study of classroom discourse: the process-product research tradition, which aims to determine what teaching practices are effective for student achievement; and descriptive classroom research, which describes the process occurring in the classroom (Cazden 1986). The present study will mainly focus on the latter through both quantitative and qualitative analyses.

This section will explore the relevant issues for this thesis concerning classroom interaction. These are, firstly, the learning opportunities interaction offers, focusing on specific interactional features, namely scaffolding and IRF patterns (with special emphasis on teachers' questions and the follow-up turn); secondly, the effect of interaction on second language learning, with a focus on the interactionist hypothesis and corrective feedback. And finally, the study of interaction in CLIL classrooms and how it differs from interaction in other types of classroom settings.

3.1.1 Classroom interaction and learning opportunities

Barnes (1975: 14) claimed that the main purpose of classrooms and education in general is communication. Therefore, it follows that communication and learning are part of the same process, that of education (Barnes 1975). Interaction and learning are, then, inextricably linked (Van Lier 1996; Vygotsky 1978). If enhancing learning is the goal of education, then an understanding of the interactions that take place in a classroom is primordial (Walsh & O'Keeffe 2010). According to Barnes (1975), different types of interaction lead to different types of learning (memorising, reasoning, creating…). For instance, learners justifying and explaining their ideas are able to better understand the subject at hand (Mercer 2000).

One type of interaction that has been acknowledged to be crucial to students' learning is *dialogic interaction* (Wells 1999; Alexander 2008). In this type of interaction, discourse is jointly constructed by teachers and students, helping the learners develop reasoning and inquiry skills, which are especially difficult in scientific subjects (Nassaji & Wells 2000; Taasoobshirazi et al. 2006). In Mortimer and Scott's (2003) terminology, this kind of interaction is called *interactive/dialogic* (see Van Lier 1996 for a previous classification). In it, teacher and students explore ideas and viewpoints, creating new meanings, listening to each other, posing questions, etc. Another characteristic of dialogic interaction is *contingency*. Contingency increases learning opportunities and depth of learning (Van Lier 1996; see also Black & Wiliam 2009), as

it relates new and known material, sets up expectancies for what comes next, validates preceding and next utterances, is never entirely predictable or unpredictable, promotes intersubjectivity, and ensures continued attention (Van Lier 1996: 193). Alexander (2004: 28) describes dialogic teaching as collective (teacher and students together), reciprocal (teacher and students listen to each other and share ideas), supportive (pupils are not afraid of wrong answers and they help each other to reach common understandings) (see also Black et al. 2003; Van Lier 1996; Harrison & Howard 2009), cumulative (teachers and children build on their own and each other's ideas and chain them into coherent lines of thinking and enquiry), and purposeful (teachers plan dialogic teaching to reach certain educational goals).

However, dialogic teaching is not practised in many classrooms. In fact, Mortimer and Scott (2003) argue that classrooms in which the teacher speaks most of the time and asks questions and students are limited to answering those questions are more common. This latter presentational mode of teaching is what Barnes called the *Transmission Model*, in which the teacher sees education as transmitting knowledge and students should be tested about that knowledge, thus being passive agents whose main task is memorize content (Barnes (1975: 144; see also Alexander 2004). On the other hand, in the *Interpretation Model*, linked to dialogic types of interaction, the teacher gives the learner an active role, emphasising context, students' reshaping of knowledge through interaction, and learners' interpretations as crucial to learning. In the Interpretation Model teachers would tend to reply instead of assess, and to negotiate instead of present. Researchers do not, however, argue that a more authoritative presentational kind of interaction should never occur in classrooms; rather, both types of interaction should alternate depending on the teachers' goals at different moments of the learning process (Barnes 1975; Van Lier 1996; Mortimer & Scott 2003; Alexander 2004; Wragg & Brown 2001). However, if the Transmission Model is always present, students run the risk of not being cognitively engaged and challenged (Alexander 2004).

There have been many studies focusing on classroom interaction in different contexts, such as L1 content classrooms or second language classrooms. Studies focusing on classroom discourse in L1 content classes have concluded that interaction is crucial for students' content learning, especially if it is scientific content (see Edwards & Mercer 1987; Lemke 1990; Barnes 1975; Duschl & Gitomer 1997; Mortimer & Scott 2003). The importance of interactive communication has also been emphasized in the context of second language classrooms through different SLA models, such as the Interaction Hypothesis (Long 1996) and the Output Hypothesis (Swain 1985). The former states that interactional modifications realized during interaction help acquisition while the latter claims that students also need to produce output in order to test hypotheses, notice learning gaps and reflect on language.

It seems, then, that interaction is a prerequisite for content and language learning, even in the L1 (Hatch 1978; Wells 1985). However, it is not easy to define learning (increased knowledge or improved performance), as it is often not observable, and it may be the result of a number of events interconnecting in complex ways (Van Lier 1988). In applied linguistics, there are different conceptualizations of learning. In cognitive SLA, learning may be conceived as something primarily individual and cognitive, involving changes of mind (Walsh & Jenks 2010; Ellis 1994). In turn, in the sociolinguistic tradition, learning is seen as a social process more than as a product (as in the SLA tradition), and learning and use are inseparable from interaction (Walsh & Jenks 2010; Gibbons 2003; Donato 2000). Some linguists advocate for the social and the cognitive interdependency and interrelationship (Batstone 2010; Ellis 2010b). In this thesis, we follow the claim that an important part of the process of learning is embodied in interaction, where moments of potential learning can be observed (Pekarek Doehler 2010; Seedhouse & Walsh 2010). This idea is related to what Seedhouse and Walsh (2010) have called *Classroom Interactional Competence* (CIC), defined as "teachers' and learners' ability to use interaction as a tool for mediating and assisting learning" (Walsh 2006: 130). Analysing classroom interaction enables researchers to identify strategies that teachers and students use to enhance learning, such as scaffolding, opportunities for learning, negotiating meaning, appropriating new ideas and concepts (Seedhouse & Walsh 2010). Co-construction of meaning and understanding is crucial for creating and maintaining spaces for learning (Seedhouse & Walsh 2010). CIC is seen as a fundamental competence for both teachers and students in CLIL classrooms, "fundamental to the very accomplishment of the CLIL lesson as an event" (Morton 2012: 86). We have to be careful, though, not to assume participation and learning are the same thing (not even participation and understanding or participation and engagement), but rather participation may affect learning because it offers space for cognitive processes and opportunities for learning (Leung 2010; Appel 2010). Nonetheless, participation presupposes attention, and attention is a requisite for learning (Van Lier 1988). In the same way, active participation is key for interaction and communication, and hence, for learning (Van Lier 1988).

Focusing now on classroom activity types, whole-class interaction is the rule in most classrooms (Lyster 2007; Dalton-Puffer 2006; Alexander 2004). This mode of interaction has been said to provide excellent opportunities for second language learning (Netten 1991; Haneda 2005; see Alexander 2004 for whole-class teaching being as strongly correlated to high standards and school success as to the opposite). However, other authors claim that more group work is desirable (Barnes 1975), as teacher-class discussions tend to relegate students to passive agents. Since group work is not always possible, Barnes (1975: 186) argues that teacher-class discussions can be changed so that students are more involved and engaged. Several

researchers, then, call for a mixture of teacher-fronted classrooms and group work sessions (Van Lier 1988; Harrison & Howard 2009). Other authors have affirmed that the quality of interaction is more important than the type of activity or group organization (Gibbons 2003; Black & Wiliam 1998a; Alexander 2004). In fact, studies dealing with quality of discourse have shown significant differences in students' learning when high quality and poor quality interactions are compared (Clarke 1988;; Pryor & Torrance 1996; Carlsen 1991).

To sum up, the importance of classroom interaction for any type of learning has been widely acknowledged (although it is not the only factor). Furthermore, the type of learning is affected by the type of interaction. Some types of interaction lead to deep learning, whereas others promote a more superficial short-term kind of learning. Two features of classroom interaction that have been explored in terms of their benefits for learning are *scaffolding* and the *Initiation-Response-Follow-up/Feedback (IRF)* pattern.

3.1.1.1 Scaffolding in interaction

Scaffolding (or *mediation*) is the term that refers to the guidance that adults provide to children in their Zone of Proximal Development (ZPD) to move their learning forward, creating supportive conditions under which novices can participate and extend their knowledge (Wood et al. 1976; Donato 1994; van de Pol et al. 2010; Gibbons 2003; Cazden 2001; Wells 1999; Poehner 2008). The teacher, or expert, is an active participant, who has to be continually revising the scaffold to keep it in accordance with learner's capabilities. If new learner capabilities emerge, then scaffold has to change (be reduced) (Donato 1994). Scaffolding is dialogically constituted, knowledge is co-constructed in a shared activity, and that allows the novice's internalization of that knowledge (Wertsch 1979a). The support is given in a variety of manners: modelling, questioning etc. (van de Pol et al. 2010). Sometimes, scaffolding can be given in the form of *other-repair* (the teacher or other classmates correct a student's mistake). Other times, it is given in such a way that *self-repair* is encouraged, that is, the student corrects him/herself (see 1.3.2 below).

Scaffolding is characterized by three main elements: contingency, fading, and transfer of responsibility (Aljaafreh & Lantolf 1994; Van Lier 1996). Scaffolding in the ZPD is graduated and contingent: firstly, responsibility is shared by expert and novice (other-regulation), and as the learner's competence, appropriation, and internalization grow, so scaffolding is gradually reduced until the learner functions independently (self-regulation) (Aljaafreh & Lantolf 1994; Lantolf & Thorne 2006). Scaffolding can never be the same in different situations (van de Pol 2010).

Scaffolding is vital if interaction is to be contingent and dynamic. Furthermore, van de Pol (2010), after doing a review of studies focusing on scaffolding, concluded that it is effective to enhance learning. However, analysing scaffolding in a systematic way is very complex due to the difficulty in its measurement (van de Pol 2010). At the same time, the unit of analysis of the studies reviewed varies greatly, and that can influence the outcomes and impede the comparison across the results of the different investigations (van de Pol 2010). Van de Pol et al. (2009) studied scaffolding and its contingency, discovering that teachers were very little contingent, not supporting and not diagnosing students' understandings. Plausible explanations offered for non-contingent teaching are time constraint, explanation without diagnosis as something deeply ingrained in teachers' minds, and teachers relying on their beliefs about what is difficult for students. If teachers teach contingently, they are less directive and they do not pursue their own agenda; rather, they adapt to students' needs (van de Pol et al. 2009). Van de Pol et al.'s (2009) work is one of the few which combine informal AfL (see section 2) and scaffolding. Their model, which was designed taking the two approaches into account, is thus very relevant for the present study.

Scaffolding provided during classroom interaction cannot be planned beforehand, as the teacher makes decisions on the fly to encourage students' participation using as a basis students' responses (Van Lier 1996). Van Lier (1996) further suggests that this type of improvized scaffolding is what makes good pedagogy and defines a good teacher. During a class, however, both the planned and the improvized need to coexist: the organized part is needed to create stability and reassuring predictability; the improvized part is key for contingent social interactions (Van Lier 1996).

3.1.1.2 IRF patterns

One of the most pervasive and ubiquitous patterns that appear in teacher-student interactions are *IRF sequences* (Sinclair & Coulthard 1975; also named *triadic dialogue* by Lemke 1990): Initiation-Response-Feedback. In this pattern, the teacher asks a question, the students respond, and then the teacher either assesses its correctness or provides feedback (Cazden 2001).

The pervasiveness of this pattern (both in primary and secondary education, and in any type of classroom: L1 content classrooms, second/foreign language classrooms, CLIL classrooms) has been described repeatedly in the literature (Barnes 1975; Mehan 1979; Sinclair & Coulthard 1975; Lemke 1990; Nassaji & Wells 2000; Lee 2007; Nikula 2007). Mercer (1995) refers to IRF sequences as the "two thirds" rule, where two thirds of the time somebody is talking (the teacher), and two thirds of teacher's talk is about lecturing or asking questions. Van Lier (1996)

calculated that over 50% of the exchanges in Sinclair and Coulthard's, Mehan's and Wells' studies fall into this interactional pattern.

Hall and Walsh (2002: 196-197; see also Mortimer & Scott 2003; Barnes 1975; Wells 1993; Van Lier 1996) distinguished between IRF and IRE (Initiation-Response-Evaluation), claiming that researchers using IRE view teaching as a process of transmission (the teacher passes information on to the students, who have to recall it), and researchers who use IRF view teaching as a process of inquiry, exploration, problem solving, and hypothesis testing (students are also responsible for their learning and this leads to higher student achievement). Either the students have to show what they have learned and the teacher assesses it (IRE), or the teacher looks for the students to be actively engaged and participating in the discussion, so that IRFs can be opened up to more "mutually contingent interactional formats" (IRF) (Van Lier 1996: 154). However, IREs have been claimed to be the norm in most classrooms, making classroom interaction unequal and asymmetrical (Applebee et al. 2003; Hardman et al. 2003).

Criticism to IRF patterns has been widely pointed out (Barnes 1975; Cazden 2001; Sinclair & Brazil 1982; Lemke 1982; Edwards & Westgate 1994; Mehan 1979; Markee & Kasper 2004). Some of this includes that teachers have the right to select topic, they have the first turn (students rarely can initiate discourse or ask their own questions), they also have the right to allocate turns, to address any student at any time, to interrupt students, to disagree with students, and to control students' behaviour (especially when they are pressured by him/her). Teachers' and students' rights are tacitly agreed, establishing an asymmetrical relationship and constraining students' opportunities to participate in the construction of learning. And students are not able to intervene with long and complex turns as well as they cannot pursue their own ideas instead of the teacher's (Van Lier 1996; see also Alexander 2004; Ruiz-Primo & Furtak 2006, 2007; Nassaji & Wells 2000; Nystr& & Gamoran 1997; Hardman et al. 2003). As a result, students' initiative, independent thinking, and creativity are hindered. In the same way, students' interpretations are not valid and they are not responsible for their learning; therefore, teaching effectiveness is compromized (Barnes 1975; Cazden 2001; Lemke 1982; Nystrand 1997; Hall & Walsh 2002; Van Lier 1988, 1996; Hardman et al. 2003). Furthermore, students' participation is limited, and participation is one of the sources for learning (Nassaji & Wells 2000; Hall & Walsh 2002; Donato 1994, 2000; Wenger 1998; Anderson et al. 2007; Hickey et al. 2006). Further criticism states that IRFs facilitate teacher control of the interaction but not student learning (Cazden 1988) Moreover, Nystrand (1997) found IRFs to be negatively correlated with learning. To sum up, then, and in Van Lier's words (Van Lier 1996: 151), "[t]he IRF structure therefore does not represent true joint construction of discourse".

However, IRFs are not necessarily always good or bad in themselves, but rather all this depends on the purpose, the goals and the occasion (Wells 1993). In addition, IRFs can lead to very different levels of student participation and engagement, depending on the purposes they serve, and on "larger goals by which those purposes are informed" (Wells 1993: 3). Van Lier (1996: 153) proposed an IRF continuum, each pattern having clear and different goals: repetition, recitation, expression, and cognition. The latter two are also referred to as *responsive teaching*. Practices close to repetition or recitation in an assessment orientation do not necessarily have to be rejected, but they have to have a purpose (see also Wells 1993; Nystrand & Gamoran 1997; Mortimer & Scott 2003). If IRFs turn to a more participation orientation, emphasising cognition and expression, and moving away from repetition and recitation, they can become more beneficial and turn towards a more contingent form of instructional interaction. Recitation teaching would be a previous stage to responsive teaching (using contingency questioning, closer to conversational dialogue) (Van Lier 1996: 160; see also Alexander 2004). Responsive teaching should be used when we want to move students ahead in their ZPD and trigger autonomy.

Some researchers have claimed that small changes in traditional IRFs can move the lesson closer to expression, and students can benefit both socially and cognitively (Cazden 2001; Gibbons 1998). Some of these changes include changing speaking rights (with students being able to select themselves for talking or addressing each other, that is, being able to initiate discourse), teachers' using fewer display questions, more recasting, and letting students take more control and responsibility in their own learning process (Gibbons 1998; Cazden 2001; Barnes 1975). As a result of these changes, student participation changes over time, that is, they progressively make more initiations and respond to teachers' questions more appropriately in both form and content, which also make classroom interaction move away from recitation and repetition (Mehan 1979). In fact, Mortimer and Scott's study (2003) showed how students took control over the agenda by raising new questions and issues at certain times, responding and giving feedback to each other, without the teacher's intervention. Teachers in their study recognize that they became more aware of their own teaching, that they changed their way of interacting with students, now valuing more students' ideas, and that they became more capable of checking their students' understanding. According to Mortimer and Scott (2003: 116), the active engagement of students, taking the initiative and participating spontaneously, was due to the fact that teachers, from the very beginning, took their ideas into account.

IRFs have also shown other positive effects. For instance, some researchers have found these patterns to be an effective way to check students' understandings, to guide students' learning, to achieve goals of education, to make educationally important knowledge salient and fundamental

for the co-construction of knowledge, to make sure that the interaction proceeds orderly, that class discussion is led to the desired direction, that all participants contribute and all participants benefit from co-construction of knowledge (Nassaji & Wells 2000; Mercer 1992; Newman et al. 1989; Seedhouse 1997; Van Lier 1996). Van Lier (1988) considers IRF patterns important for the teacher to control the lesson, and that is not an unimportant issue. In addition, IRF sequences are not unnatural, for they also occur in adult-child interaction (Seedhouse 1997).

Finally, IRF sequences have been found to be different depending on the subject. Nassaji and Wells (2000), who specifically compared Science and Arts (– Literacy and History –), discovered that Science classrooms used a wider range of episodes. Also, in Arts, the percentage of factual questions was lower than in Science and the percentage of negotiatory questions was higher. Student initiations were more frequent in Arts, too (Nassaji & Wells 2000: 389). In the same way, differences are expected in this study when IRF patterns are compared in different subjects (Arts, Science, Citizenship, and Drama). The sections below will focus on two well-researched parts of the IRF exchange: teachers' initiations (through questions) and teachers' follow-up turns.

3.1.1.2.1 Teachers' questions

> No one can teach, if by teaching we mean the transmission of knowledge, in any mechanical fashion, from one person to another. The most that can be done is that one person who is more knowledgeable than another can, by asking a series of questions, stimulate the other to think, and so cause him to learn for himself
>
> (Socrates, 5th century BC, in Wragg & Brown 2001: 27).

Teachers' questions have been the focus of extended research, given their frequency, their pedagogical goal, and the control they exert over talk and curriculum (Cazden 1986: 440). Among their main functions, they serve to orient children, to establish shared knowledge, to find out what the learner already knows (to teach them accordingly), to stimulate recall, to encourage problem-solving, to deepen understanding, to manage the classroom, to revize, to check understanding, to encourage a more active engagement in learning, to offer practice on the material, and to provide feedback that clarifies understanding and corrects misconceptions (Cazden 1986; Crooks 1988; Wragg & Brown 2001). In sum, teachers' questions facilitate learning (Wragg & Brown 2001). In fact, some studies have reported that the frequency of teacher questioning correlates with student achievement (Crooks 1988). Therefore, some authors have claimed that questioning is at the heart of effective teaching because it gives students the opportunity of improving their thinking and learning: good questioning leads to

successful and interactive teaching, bad questioning can lead to confusion and misunderstanding (Wragg & Brown 2001).

There have been different categorizations of teachers' questions, the best known being the differentiation between open/closed questions (Barnes 1969) and the display/referential distinction (Mehan 1979). The first classification (open/closed questions) can pose more problems than it may seem prima facie. A closed question in one context can be open in a different one (Wragg & Brown 2001). In the same way, we can find pseudo open questions, that is, open in form but closed in function, when the teacher is looking for one specific and single answer (Barnes 1969; Wragg & Brown 2001). In his data, Barnes (1969) discovered that more than 50% of the pseudo open questions did not receive an answer from the pupils, so the teachers had to answer the questions themselves. Similarly, closed questions usually receive short answers, inhibiting thinking and discussion (Barnes 1969; Wragg & Brown 2001; Dalton-Puffer 2007).

As for the second categorization (display and referential questions), asking learners questions whose answer is known by the teacher (display questions) has three different functions: move the lesson forward as planned; help students accomplish an academic task; and help teachers assess student learning (Cazden 1986). In turn, referential questions (those questions whose answer is not known by the teacher) are said to be more desirable than display ones, as the latter limit students' possibilities to express their own ideas and the responses they get are shorter (Long & Sato 1983; Romero & Llinares 2001; Dalton-Puffer 2007; Lyster 2007; Pascual Peña 2010). They are also considered by some as more effective, as some studies have proved that students can guess the correct response to display questions just from learning the patterns of communication and not from actual content learning (Lundgren 1977). However, there have been researchers who have claimed that display questions are important and effective for content learning and to verify content comprehension in bilingual classrooms, and that these are as effective at eliciting students' extended responses as referential questions (Musumeci 1996; Salomone 1992). Lyster (2007) argues that criticism of display questions may apply more to EFL classrooms than to content-based classrooms (but see McCormick & Donato 2000), for despite the constructivist co-construction of knowledge and negotiation of content, teachers need to be responsible and scaffold novices in a dialogical way. To do that, asking only questions to which they do not know the answer would not be desirable. In sum, what seems important is that teachers display a variety of question techniques, including both display and referential (Lyster 2007). In fact, Haneda (2005) observed that both display and referential questions had the same effectiveness.

Categorising teachers' questions according to a cognitive scale has been a common method for analysing classroom discourse (Cazden 1986). The most influential one has been Bloom's *Taxonomy of Educational Objectives* (1956). This taxonomy classifies thinking according to six cognitive levels of complexity. These six levels are, from lowest to highest: knowledge, comprehension, application, analysis, synthesis, and evaluation (the original terminology has changed in later revisions of the taxonomy) (Krathwohl 2002). Bloom's taxonomy is hierarchical, that is, mastery of a level presupposes mastery of lower levels (Krathwohl 2002). Especifically in CLIL, Dalton-Puffer (2007) followed this cognitive approach to discourse to develop another classification of questions for CLIL classrooms: questions for facts, explanations, reasons, opinions, and questions asking for meta-cognitive information. Based on this classification, studies in different CLIL contexts have shown that teachers mainly ask questions for facts (Dalton-Puffer 2007 in Austria; Pascual Peña 2010 and Llinares and Pascual 2014 in Spain; see also Menegale 2011 in Italy & Schuitemaker-King 2012 in Holland) and that these questions are followed by short and simple responses from the students. In turn, these studies have shown that questions which seek students' opinions as well as their metacognitive processes (that is, higher-order and authentic questions) trigger a more dialogic form of discourse and a more exploratory stance of the topic under discussion, which lead to more extended students' responses and deeper learning (Dalton-Puffer 2007; Alexander 2004; Nystrand 1997; Nassaji & Wells 2000; Mercer 1995; Nystrand & Gamoran 1997; Wragg & Brown 2001). Unfortunately, these types of questions are very infrequent in any classroom context (Wragg & Brown 2001), including CLIL (Dalton-Puffer 2007; Pascual Peña 2010). A possible explanation, at least in the case of the Spanish context, may be the pedagogical culture of certain subjects like History, in which factual content is given more emphasis than argumentation or problem-solving skills (Llinares & Pascual 2014). Opposite findings were shown by Nassaji and Wells (2000) in Primary L1 classrooms, who found factual questions to be less frequent than negotiatory questions (questions to be answered through negotiation and exploration, not by mere display of information). Be it as it may, Harlen and Winter (2004) claim that changing a low-order question into a high-order one is very easy: it is a matter of students responding what they think, not what the teacher is looking for in the form of a correct answer. For instance, a question like "why do we eat food?" might be changed to "why do you think we eat food?" (Harlen & Winter 2004: 398). Teachers should ask students for their reasoning in order to avoid rote memorization and foster deep learning, making sure they understand what they are saying and the meanings they are constructing (Mohan et al. 2010; see also Dalton-Puffer 2007).

In general, different reviews on teacher questioning seem to conclude that higher level questions normally foster student achievement (or at least they do not harm it), they enhance learning, retention, transfer, interest, and development of thinking skills (Winne 1979; Crooks 1988: 455). However, guidance and training is needed for students to be able to answer this type of questions and they need to be used consistently over different educational levels (Crooks 1988). All in all, what seems to be important in teacher questioning is variation through different types of questions and letting students ask questions too (Wragg & Brown 2001; Lyster 2007; Brophy & Good 1986). This variation in questioning will probably lead to variation in different types of interaction, all of which have their space and their purpose in the classroom context. Within these types of interaction, dialogic teaching is especially encouraged, where questions often provoke thoughtful answers and answers provoke further questions. In the same way, teachers both encourage participation and extend students' understanding through their questions. Students are also encouraged to initiate discourse (ask questions) and provide explanations too. In dialogic teaching, students are not worried about guessing the right answer; on the contrary, students' responses to questions move beyond yes/no answers (or simple recall responses) to extended answers involving reasoning, thinking, and hypothesising (Alexander 2004: 43).

3.1.1.2.2 Teachers' follow-up

The follow-up turn, that is, how the teacher uptakes a student's response has been the focus of a lot of research, too, and its complexity and importance have been highlighted in numerous studies (Hall 1998; Nassaji & Wells 2000; Wells 1993; Young 1992; Jarvis & Robinson 1997; Zhang Waring 2008; Nystrand & Gamoran 1991; Lee 2007; Barnes 1975; Carlsen 1991; Cazden 1986, 1988; Mortimer & Scott 2003; Lyster 2007; Alexander 2004; Haneda 2005). A child's answer can never be (as they tend to be) the end of an exchange: as important as questions are answers and what teachers do with them (Alexander 2004).

Follow-up moves can include *evaluative feedback* (IRE) or *non-evaluative feedback* (IRF) (Barnes 1975; Wells 1993; Mortimer & Scott 2003). The former has been related to a Transmission Model of teaching, and the latter to an Interpretation Model of teaching, or teaching as a process of inquiry. Also, certain actions that teachers perform in third turns (elicitation, repetition of a student's answer, praise, and reformulation) can become formulaic and automatic, and therefore, lose the force of stimulating and advancing students' thinking (Mercer 2000; Alexander 2004).

There are a number of characteristics that make follow-up turns particularly important. Follow-up turns are contingent upon the second student turn, and hence, not predictable but responsive

(Lee 2007). Follow-up moves are also retroactively contextualizing, they recontextualize student's response structurally or thematically. Thematic recontextualization occurs frequently, when the teacher alters or enriches the content of the response, placing it in a wider and more relevant context for the thematic aims of the lesson (Lee 2007; Alexander 2004; Barnes 1975). Another characteristic of third turns is that they show multiple levels of meaning-making, even more so in a content and language integrated classrooms (Lemke 1982). In addition, third turns are expected to go beyond evaluation of the student's response (non-evaluative feedback rather than evaluative feedback), which is critical for the relationship between classroom discourse and student learning (Nystrand et al. 1997; Haneda 2005). In dialogic teaching, feedback on responses provides informative diagnostic feedback rather than a positive or negative comment or a repetition of the student's answer; uses reformulation in a way that avoids ambiguity regarding the approval or disapproval of the answer; uses praise appropriately; keeps lines of enquiry open rather than closing them; encourages learners to articulate their ideas; uses students' responses to involve them in the learning discourse (Alexander 2004). In the same way, third turns are important to sequence and structure lessons, change topic, move on the discussion, and make students feel their contributions are relevant (Alexander 2004). Teachers can make students' contributions relevant in different ways, such as building or inviting other students to build on the response, linking previous students' contributions with new ones, or incorporating students' contributions into summaries and reviews (Alexander 2004). As well as the type of question, whether the teacher evaluates the answer or not has an effect on the students' contribution: when there is no evaluation, students' contributions are longer and more complex, and when there is teacher evaluation, students' responses are short and simple (Nassaji & Wells 2000: 400-401).

Non-evaluative feedback moves have been reported to create more opportunities for learning than evaluative ones (Wells 1993; Barnes 1975; Mortimer & Scott 2003; Nassaji & Wells 2000; Hall 1998 in Spanish as a FL class; Nystrand 1997 in language art classes; Rex & McEachen 1999 in English literature classrooms; Cortés-Conde 2000, and Boyd & Maloof 2000 in university English as a second language classrooms; Consolo 2000 in Brazilian EFL classes; Duff 2000 in EFL classrooms in Hungary; Sullivan 2000 in Vietnamese EFL classes). Of all these studies, special attention will be given to Wells (1993), Barnes (1975), Mortimer and Scott (2003), and Nassaji and Wells (2000), since they are the most relevant ones for this study.

If the third turn is used to other than evaluate, then IRF sequences can turn into IRF-R-F, that is, the feedback move is followed by another student response, which again receives elaborative feedback and so on (Mortimer & Scott 2003). In this line, Wells (1993) found that when teachers did not evaluate in the third turn, students engaged in long interventions. He therefore

distinguished between evaluative third moves and those that extend students' answers, make connections, mark their importance, etc. As Barnes said (1975: 111), these different functions have different consequences for students' learning (for instance, if the teacher assesses, exploratory talk will not occur). Wells (1993: 35) argues that if the third move is not used for evaluating but rather as a non-evaluative feedback move, meaning can be co-constructed and it can be the point of departure for a new learning cycle. In this way, teachers challenge students to extend thinking and engagement and they provide chances for students to take ownership of the ideas (Nystrand 1997). Nassaji and Wells (2000) show how IRFs can take a variety of forms and functions, depending on the goal of the activity and on how teachers use the third move. Therefore, IRF patterns can match different teaching philosophies and so they can be present even when more dialogic style of teaching is trying to be practiced (Nassaji & Wells 2000). Nassaji and Wells (2000) claimed that, even when the initiating question is a display question, the dialogue gets more equal if the teacher avoids evaluation. They also discussed how frequently teachers take up different functions of the third move and what implications it has for participation. Among their findings, they showed how teachers, when making questions for opinions, still evaluated the response, which would not be in consonance with an environment of inquiry. They also found out that negotiatory questions gave birth to other types of follow-up moves, not evaluations, whereas factual questions led more frequently to evaluative feedback. The explanation to this can be that the type of evaluation was different: for factual questions, teachers evaluated their correctness; for negotiatory ones, teachers evaluated (praised) the relevance of the contribution, repeated the response for everybody to hear or anticipated further elaboration (Nassaji & Wells 2000: 397). To sum up, these authors claim that follow up moves allow the teacher to work with the students' responses in a variety of ways.

Finally, the same follow-up move can serve different purposes at the same time. For instance, third turns can evaluate the previous second turn and initiate another three-turn sequence (Van Lier 1988). Similarly, Wells (1993) explained how the third move can be a question, which, at the same time, would be the initiating move of the next exchange and would encourage the learner to take responsibility and make what s/he said in the previous utterance more comprehensive. This resulted in students' contributions being longer and more complete.

After considering the importance of interaction in education, the next section will deal with the issue of how interaction can enhance second language learning, which is one of the objectives of CLIL. To do this, Second Language Acquisition (SLA) literature on interaction will be reviewed, specifically the *interaction hypothesis* and *corrective feedback*.

3.1.2 Classroom interaction and second language learning

As mentioned above, interaction in education is a key feature for learning, even more so in a second language classroom. One of the most influential approaches in the study of interaction in second language acquisition (SLA) research is the *Interaction Hypothesis*. This section will explore this approach, paying special attention to negotiation of meaning. Likewise, *feedback* (corrective feedback) and its relationship with L2 learning will be thoroughly scrutinized, as it is one of the main factors influencing L2 learning during interaction and an important aspct in the present study.

3.1.2.1 The Interaction Hypothesis (Long 1996)

Second language classrooms have been extensively researched. Specifically within the cognitive paradigm, which has dominated SLA, the model of language acquisition is the input-output model: student receives input from the environment, processes it and produces output (Gass & Selinker 2001). Negotiation of meaning is key to provide learners with learning opportunities (Long 1985; Pica et al. 1989). This negotiation allows the learner to notice the gap between the target language and their interlanguage, to receive input of higher comprehensibility, and to produce modified output (Gass & Mackey 2006). Thus, those tasks which involve more negotiation of meaning will provide greater opportunities for learning.

Input is the language available to learners through any medium: they need this linguistic evidence to formulate and test hypotheses about the second language (Gass & Mackey 2006: 2; Krashen 1985). *Comprehensible input* is necessary for SLA (Krashen 1985). There are different types of input (Pica 1994): positive input, providing learners with models of the TL; enhanced input, making salient specific features to assist learning; negative input, to provide learners with metalinguistic information about their output and their IL, noticing non-target-like utterances. Many researchers have claimed that it is difficult to connect comprehension of meaning and internalization of L2 forms (e.g., Chaudron 1985; Gass 1988; Sharwood Smith 1987;). They contradict Krashen (1980, 1983, 1985), who thinks that comprehension of meaning suffices to have access to L2 forms. Attention to L2 form is another crucial feature in SLA (Long 1990; Schmidt 1990). There are some L2 structures which are imperceptible in the input or which overlap with L1 structures (see Lightbown & Spada 1990). Therefore, attention to L2 form is key if input is to have a positive effect in L2 learning.

Output is the language that learners produce (Swain 1985). According to Swain (1985), modified and comprehensible output is necessary for second language acquisition to occur. Therefore:

> learners need to be pushed to make use of their resources; they need to have their linguistic abilities stretched to their fullest; they need to reflect on their output and consider ways of modifying it to enhance comprehensibility, appropriateness, and accuracy (Swain 1993: 160-161).

Modified output (after feedback) is said to promote learning because it promotes learner's reflection on his/her language, pushes learners to produce more accurate, complex, and comprehensible output (Swain 1985, 2005; Swain & Lapkin 1995), and provides learners with opportunities for hypothesis testing (Swain 1985, 1995, 2001). In addition, output will also help learners notice what they do not know or know partially (Swain 1995; 2000, 2001). Swain (1985) proposed the *Output Hypothesis* after seeing that competence in spoken production in Canadian immersion programmes was way below listening comprehension. In 2001, Swain (2001) reconceptualized output as dialogue serving cognitive and communicative purposes (Swain 2001).

The third element in the interactionist model is *feedback*. To be effective, feedback has to be developmentally helpful (Gass & Mackey 2006). Learners have to make connections between their erroneous utterances, feedback, and output. If the feedback does not provide the right utterance, students need to be in the appropriate developmental level which allows them to figure out how to make the correction. Pushing learners in their output, promoting self-correction, rather than giving them the right answers, benefits their interlanguage development (Swain 1985; Van Lier 1988; Ellis 1997). Pica et al. (1989) found that modified output took place mainly when NS signalled the need for clarification rather than when providing a correct model themselves or a recast. In the same line, Nobuyoshi and Ellis (1993) discovered short-term benefits for students receiving clarification requests. However, there is still a debate about whether modified output is indeed necessary or not (Lyster 1998b). Schachter (1983) and Long (1977) warn about the fact that imitation or expression of alternative correct ways of saying something after teacher feedback does not mean learning, let al.one long-term learning. Also, Gass (1988) stated that providing learners with explicit language information does not mean they will convert it into output. She also argues that frequent negative evidence is needed for students to notice discrepancies between their interlanguage and the target language, and to prevent fossilization from happening. In immersion contexts, and extensively, in CLIL classrooms, however, there are two main foci, content and language. That is why learners may

have more difficulty in noticing these mismatches, and thus, signals facilitating repair are more needed than just providing target forms for students to be able to notice the gaps (Lyster 1998b).

The combination of these three elements (input, output and feedback) constitutes the Interaction Hypothesis. Basically, it states that

> through input and interaction with interlocutors, language learners have opportunities to notice differences between their own formulations of the target language and the language of their conversational partners. They also receive feedback which both modifies the linguistic input they receive and pushes them to modify their output during conversation (Gass & Mackey 2006: 1).

During interaction, there are moments in which lack of understanding occurs and negotiation of meaning needs to be present. During this negotiation, input may be modified, learners are given feedback on their production, normally for them to notice gaps between their output and the target language and hence be able to modify their output (Gass & Mackey 2006; Long 1996; Pica 1994). Interaction, then, is essential and has positive effects on L2 development (Long 1996; Gass & Varonis 1994; Ellis et al. 1994; Pica et al. 1989; Chaudron 1986; Doughty 1988, 1992).

It has been proven, then, that negotiation helps L2 learning; however, it does not account for all learning (Pica 1994), learner factors can also have an influence (Pieneman 1989). In addition, negotiation seems to work more on lexical and larger syntactic units, but rarely on morphology (Pica 1994: 22; Pica, Kanagy & Falodun 1993). Besides, learners and interlocutors can communicate messages through interaction, but not necessarily with target-like forms (Pica 1994: 22; Sato 1986).

Although negotiation of meaning has been said to be broadly present in content-based classrooms (Genesee 1987; Met 1994), it has not yet been proven that student-student interactions and teacher-student interactions in those settings lead, per se, to second language learning and development (Lyster 2007: 105; Aston 1986; Lyster 2002; Musumeci 1996; Foster & Ohta 2005). In content-based classrooms, teachers interpret students' speech instead of encouraging negotiation for learners to modify their output when necessary (Musumeci 1996; Harley 1993). Negotiation of meaning focuses on meaning and not on form, on form only to the extent that this is needed for comprehensibility (Pica 1994). Lyster (2002, 2007: 106) emphasizes that teachers become experts at understanding students' interlanguage, and that can be a disadvantage for language development and accuracy, because once students have the sufficient language repertoire to meet their communicative needs, negotiation of meaning becomes limited. Thus, apart from negotiation of meaning, negotiation of form is also important

for second language learning (Lyster 2007; Swain 1995). A counterbalanced approach is advocated in immersion settings (Lyster 2007) as well as in CLIL classrooms, although in the latter case it is referred to as integration (Llinares, Morton & Whittaker 2012).

After presenting the Interaction Hypothesis, the next subsection will expand on one of its elements, feedback, due to its relevance for the present study.

3.1.2.2 Corrective Feedback and SLA

While positive evidence is the input which models the target language that learners receive (Morris 2002), negative evidence (normally called feedback, or corrective feedback) is information that a particular utterance is not correct according to target language rules (Gass & Mackey 2006; Long 1996). Negative feedback has been shown to facilitate L2 learning (Swain 1985, 1995; Long 1996; Schmidt 1990; Panova & Lyster 2002; Lyster & Ranta 1997). Corrective feedback types range in an explicit-implicit continuum. At the most explicit end, there is explicit correction, in which the teacher explicitly states there has been an error and provides the correct form (Lyster & Ranta 1997; Gass & Mackey 2006). Implicit negative feedback is more covert, the learner identifies the error and is able to correct it not because of explicit feedback but because of negotiation (Gass & Mackey 2006). Confirmation checks ("any expressions… immediately following an utterance by the interlocutor which are designed to elicit confirmation that the utterance has been correctly heard or understood by the speaker", Long 1983: 137), clarification requests ("any expression… designed to elicit clarification of the interlocutor's preceding utterance(s)", Long 1983: 137), and comprehension checks (an attempt "to anticipate and prevent a breakdown in communication", Long 1983: 136) are implicit forms of feedback (Gass & Mackey 2006: 6; Long 1996).

Teachers' extensive use of feedback is well documented (Doughty 1994; Fanselow 1977; Lyster & Ranta 1997; Roberts 1995; Llinares & Lyster 2014), and so is its ambiguity (Lyster 1998a; Allwright 1975; Chaudron 1977; Netten 1991; Allen, Swain, Harley & Cummins 1990). One of the best-known taxonomies on teachers' feedback is Lyster and Ranta's (1997). They distinguished six main feedback moves in immersion classrooms:

1. *Explicit correction*: teacher supplies the correct form and clearly indicates that what the student had said was incorrect;
2. *Recasts*: teacher implicitly reformulates all or part of the student's utterance;
3. *Elicitation*: teacher directly elicits a reformulation from students by asking questions such as "*Comment ça s'appelle?*" or "*How do we say that in French?*" or by pausing to

allow students to complete the teacher's utterance, or by asking students to reformulate their utterance;

4. *Metalinguistic clues*: teacher provides comments, information, or questions related to the well-formedness of the student's utterance such as "*Ça ne se dit pas en français*" or "*C'est masculin?*";
5. *Clarification requests*: teacher uses phrases such as "*Pardon?*" and "*I don't understand*";
6. *Repetition*: teacher repeats the student's ill-formed utterance, adjusting intonation to highlight the error.

Lyster and Ranta's (1997) clarification requests, repetitions, metalinguistic feedback, and elicitation were termed *prompts* and they imply negotiation of form when followed by student self-repair (Lyster & Ranta 1997; Lyster 2007). Different feedback types can have different effects on learning (Lyster 1998b). Prompts, for instance, unlike explicit correction, may trigger peer- or self- repair without providing the correct form (Morris 2002: 3). In the same way, prompts lead to high rates of student uptake (student's response after teacher feedback) and student repair (Lyster & Ranta 1997).

Recasts, defined as the reformulation of an error or errors partially or fully, are the most frequent type of feedback in a variety of instructional settings (EFL, ESL, immersion, content-based, CLIL), countries (USA, Australia, Canada, Korea, Japan, New Zealand, Spain), and educational levels (elementary and university levels) (Lyster & Ranta 1997; Doughty 1994; Fanselow 1977; Roberts 1995; Mori 2002; Sheen 2004, 2006; Ellis et al. 2006; Ellis et al. 2001a, 2001b; Panova & Lyster 2002; Chaudron 1977; Fanselow 1977; Doughty 1994; Lyster & Mori 2006; Lee 2007; Llinares & Lyster 2014). Research on recasts has focused on the following aspects: a) whether they contribute to learning (see Han 2002 & Long et al. 1998 for proof that learners develop metalinguistic awareness as a result of exposure to recasts; Ortega & Long 1997; Mackey & Philip 1998; Mackey et al. 2000; Leeman 2000, 2003); b) whether they lead to learner uptake, which may depend on the recast type (Lyster & Ranta 1997; Sheen 2004, 2006; Mori 2002; Chaudron 1977); c) whether recasts provide positive or negative evidence (Leeman 2003); d) the extent to which recasts are noticed by learners, which seems to depend on students' proficiency level, length, and number of changes of the recast (Mackey et al. 2000; Philip 2003); e) the relationship between recasts, uptake and L2 development (Mackey & Philip 1998; Loewen 2005). One problem with the research carried out on recasts is that the definitions of recast are not the same in all the studies, and so results are not comparable (Lyster 1998b, 2002).

As other types of corrective feedback, recasts can also be implicit or explicit. Lyster (1998a, 1998b; 2007) argues that the more explicit recasts are, the more beneficial they will be for SLA. If recasts are implicit, then they fulfil the same function as *non-corrective repetition* (Lyster 2007). Students may simply take it as an alternative way of saying something, without implying that their previous utterance was wrong. Besides, recasts do not lead to uptake or to student-generated repair because they already provide learners with correct forms (Lyster & Ranta 1997). Lyster further argues that the primary function of recasts might not be corrective (especially when they approve the content of the ill-formed utterance) and that they have more in common with *non-corrective feedback and topic continuation moves* than with corrective feedback. In this way, recasts can be perceived by learners as positive instead of negative evidence (Lyster 1998a: 71). In contrast, Long (1996) and Long and Robinson (1998) considered that recasts were effective because they were implicit, because they can focus both on form and meaning, without interrupting the flow of conversation, and allow learners to compare IL and TL forms (explicit correction interrupts the flow of the lesson and will not assist form-function mapping and so L2 acquisition is not likely to occur). Other authors prefer to think about recasts in an explicit-implicit continuum, stating that, the more explicit recasts are, the more potentially salient they will be and the more effective they are likely to be (Sheen 2004, 2006; see also Ellis & Sheen 2006). In this explicit-implicit continuum, Sheen and Ellis (2011) distinguish between *conversational* and *didactic recasts*, the former implicit and the latter explicit. Conversational recasts implicitly reformulate a student utterance in order to solve a communication breakdown, whereas didactic recasts reformulate a student utterance with the purpose of correction, since there is no communication problem. This distinction is based at the same time on Lyster and Ranta (1997), who distinguished between two functions of negotiation: conversational, which entails negotiation of meaning, to repair breakdown communication and achieve mutual comprehension; and didactic, which includes negotiation of form or the provision of corrective feedback for the student to self-repair, focusing on accuracy rather than comprehensibility. In a similar vein, Mohan et al. (2010) distinguished between formal recasts, focusing on form, and functional ones, focusing on meaning and with the intention of keeping the flow of the conversation rather than with the intention of correction. In any case, it is important to take into consideration the fact that the concept of explicitness itself may depend on external variables, such as the instructional context and the communicative orientation (Llinares & Lyster 2014).

Regarding whether recasts lead to student uptake, it is important to refer to noticing. It cannot be affirmed that no uptake implies no noticing (Sheen 2006; Ellis & Sheen 2006). However, uptake or repair is evidence of noticing and this is potential for language learning, although a

direct link between noticing and learning cannot be proved (Mackey 2006; Sheen 2006; Ellis & Sheen 2006). Different studies have investigated what types of recasts are more likely to be noticed, and hence, to be followed by student uptake. Lyster (1998a, b) discovered that recasts involving reduction of learner utterance were the ones followed by a higher rate of repair. The linguistic focus of recasts might be an important factor for their noticing, too. Lyster (1998b) and Mackey et al. (2000) show how students noticed more phonological and lexical recasts – or any other type of feedback – than grammatical ones. Philip (2003) suggested that there are different factors intervening in noticing of recasts, such as learner's proficiency, length of the recast, and number of changes made in the recast. Sheen (2004) suggested that the saliency and explicitness of recasts influence on learner uptake/repair. In this way, Sheen (2006) found that six characteristics of recasts were related to learner's uptake: mode, length, type of change, linguistic focus, and reduction. Among them, the most salient features of recasts are pronunciation-focus (in line with Mackey et al. (2000) finding that phonological recasts are more noticed by learners, and Lyster (1998b), in which study phonological recasts were followed by a high rate of uptake), and shorter length (Lyster 1998a argued that reduced and short recasts are easier for students to compare them with their erroneous utterance) and substitution.

Regarding to what extent recasts lead to higher or lower rates of student uptake when compared to prompts, some studies show prompts to promote higher rates of student uptake (Lyster & Ranta 1997; Mori 2002). Clarification requests and repetition of error could also be said to negotiate for meaning, but Lyster (2007: 108) argues they negotiate form because they push students to modify their non-target output, and because they are not uttered because teachers do not understand but because they feign comprehensibility and want to draw attention to non-target forms on purpose. Lyster's (2007) research contradicts claims that oral feedback on accuracy breaks the flow of communication (Krashen 1994) and that prompts interfere with the delivery of content (Long 2007). There are other studies demonstrating that prompts have a significant effect on IL development, both in the case of adults and children and in a variety of instructional settings (content-based ESL, ESL, communicative ESL, French immersion, EFL) (Ellis, Loewen & Erlam 2006; Havranek & Cesnik 2001; Doughty & Varela 1998; Lyster 2004a; Ammar & Spada 2006). Prompts helped learners to retrieve target-like forms not only at that time but also during subsequent processing, as opposed to students that just heard recasts of those forms (Lyster 2007: 121). Learners better remember information that they have been active in producing (Lyster 2007: 122). Students receiving prompts developed more *feedback appreciation*, that is, conscious awareness of feedback, which leads to longer-term changes (Skehan 1998; Lyster 2007: 122). Both recasts and prompts should be used in a balanced way.

Recasts will be more beneficial when target forms are beyond students' current abilities. Beyond that point, students will benefit more from prompts which push them to modify their output (Nicholas et al. 2001; Lyster 2007: 123), both because of the ambiguity of recasts and because learners have reached a developmental plateau in their use of non-target forms. However, other studies have shown opposite results regarding recasts. For instance, Ellis et al. (2001b) found that recasts were followed by the highest amount of uptake in adult ESL classrooms in New Zealand. Also, in Oliver's study (1995), recasts are followed by uptake in a higher rate due to the methodological change she incorporates, which is the exclusion of the turns in which learners do not have the chance of incorporating the recast. In Llinares and Lyster (2014), recasts are also found to be followed by uptake and even repair at high rates in CLIL and in Japanese Immersion (JI) programmes in the US, as opposed to the context of FI. These results are explained on the basis of the explicitness/implicitness of the recast: recasts in FI tend to be more conversational and, therefore, implicit; on the other hand, recasts in CLIL and JI are usually more didactic and, hence, explicit. Although recasts do not always elicit an immediate response from the learner or lead to modified output, it does not mean that they do not have an effect in the long run, regardless developmental level (Brock, Crookes, Day & Long 1986; Mackey & Philip 1998).

To sum up, a lot of research has investigated the relationship between interactional feedback (recasts and negotiation moves, according to Mackey 2000) and L2 development, usually with positive results (e.g. Ellis et al. 1994; Mackey & Philp 1998; Ellis & He 1999; Mackey 1999; Silver 2000; Mackey & Oliver 2002; Iwashita 2003; Leeman 2003; Philip 2003; Ishida 2004; Mackey & Silver 2005; McDonough 2005; Mackey in press; Oliver 2000; Ellis et al. 2001a, 2001b; Mackey et al. 2003). Each study has examined different instructional settings and thus results are not generalizable, although they have highlighted how corrective feedback may influence students' interlanguage systems. Nonetheless, the effectiveness of corrective feedback must be proven to be sustained over time (Mackey et al. 2003; Lightbown 2000).

The next section focuses on research on interaction in CLIL classrooms. It will explore what interaction in CLIL classes is like, why it is especially important, and whether it is different from interaction in other types of classes.

3.1.3 Classroom interaction in CLIL

In content and language integrated learning (CLIL) classrooms, development of curriculum knowledge and linguistic development have to go hand in hand (Llinares et al. 2012). Like in any other type of classroom, interaction is a fundamental piece in the learning process: it is an

opportunity for both language development and content learning (Vygotsky 1978; Moore 2011). Research on interaction in CLIL has focused on several aspects, such as codeswitching (Dalton-Puffer 2007; Nikula 2007a), IRF sequences and scaffolding (Dalton-Puffer 2007; Nikula 2007b; Evnitskaya 2012), repair (Dalton-Puffer 2007; Serra 2007;), pragmatics (Dalton-Puffer 2007; Dalton-Puffer & Nikula 2006; Nikula 2005, 2008; Maillat & Serra 2009), interaction in relation to genres, registers and the integration of content and language (Llinares & Morton 2010; Llinares, Morton & Whittaker 2012), and learners' oral competence (Moore 2011; Mewald 2007). General conclusions emphasize that CLIL learners have good interactive skills (better than their mainstream peers) and they speak the foreign language fluently.

Most CLIL studies on interaction have focused on teacher-led lessons and whole-class interaction at the secondary school level, as these are the most frequent scenarios in CLIL classrooms (Nikula, Dalton-Puffer & Llinares 2013). Fewer studies have focused on interaction at primary level, which is the focus of the present study. The fact that CLIL in Spain, and more specifically in the Madrid Autonomous Community, is implemented first at the primary level makes investigation in this educational period very necessary.

Despite the fact that most CLIL research on interaction has focused on whole-class sessions, there have also been studies in CLIL analysing the interaction that occurs in small group work or pair work (Bonnet 2004; Gassner & Maillat 2006; Maillat 2010; Pekarek Doehler & Ziegler 2007; Nikula 2012; Llinares & Morton 2012; Pastrana 2010; Llinares, Morton & Whittaker 2012; Llinares & Pastrana 2013). These studies prove that this type of activity gives students the opportunity to increase the quantity and the functional scope of their output, and to display a wider range of functions and registers, as well as the opportunity for students to intervene in all the three moves of IRF sequences, acting like authors and principals and not just animators of knowledge (Goffman 1981).

Specifically referring to IRF patterns, findings in CLIL are inconclusive: while Nikula (2005) found CLIL classes to have space for more dialogic teaching and less controlled interaction, Dalton-Puffer (2007) found IRFs to be restrictive and extensive in her Austrian data, with a preponderance of display questions. She argues that the communicative competence that learners can acquire and practice in CLIL classrooms is limited, as the pervasive use of IRF patterns limited the use of language functions and language opportunities were lost (Dalton-Puffer 2007). Facts were found to be the dominant landscape in Austrian CLIL classes (Dalton-Puffer 2007). This type of findings, together with the observation that the types of activities carried out in CLIL classes do not differ from those in L1 content lessons, that is, both types of classes are mainly characterized by teacher-led whole-class discussions (Badertscher & Bieri

2009), undermine claims that CLIL pedagogy is more student-centred (Mehisto, Marsh & Frigols 2008), as reported in Nikula, Dalton-Puffer and Llinares (2013: 79). In CLIL classrooms, like in most classrooms, then, teachers' control of the interaction is the norm.

Two studies have compared IRFs in CLIL and EFL classes: Nikula (2007) in Finland and Schindelegger (2010) in Austria. Their findings present a different picture in each country: IRF patterns are more frequent in CLIL classrooms in the case of the Austrian data and more frequent in EFL lessons in the case of the Finnish data. However, at a more qualitative level, they describe a similar situation, namely that IRFs are tighter in EFL classrooms. In CLIL lessons, students' responses are longer than in EFL classes, more opportunities for a varied language use in their responses emerge, and teachers use the third turn not just to evaluate the student's response but also to offer reasonings and further explanations, in extended IRFRF sequences that did not occur in EFL lessons. This fact also contributed to the sense of interactional symmetry (Nikula, Dalton-Puffer & Llinares 2013). Different CLIL classroom pedagogies (in Austria and Finland) should be taken into account for the interpretation of these results and more comparative studies across contexts are needed to explain differences and similarities in IRF exchanges.

Other findings (Nikula 2005, 2008; Moore 2011) comparing CLIL and EFL reveal that interaction is more interpersonally detached in EFL, as talk is mainly concerned with the textbook, whereas in CLIL it is more personal, as students focus on the here-and-now of the task and produce more collaborative talk. One of the reasons why CLIL students seem to engage in L2 interaction more than EFL students may be that their foreign language skills are not constantly under evaluation (Nikula, Dalton-Puffer & Llinares 2013).

As Llinares, Morton and Whittaker (2012) argue, it is not the IRF pattern itself which is restrictive or beneficial, but rather the type of activity in which it is used and the roles of the participants in it. Therefore, opportunities for participation can be very different not only when comparing CLIL and EFL, but also across CLIL subjects or even the same subject, depending on the type of activity carried out (Evnitskaya & Morton 2011).

CLIL classroom interaction research is an illustration of the shift in SLA research towards a more socially-situated approach (Pekarek Doehler & Ziegler 2007; Evnitskaya & Morton 2011; Kupetz 2011; Morton 2012). This approach does not separate content and language as two different things, rather it recognizes they are intertwined and cannot be separated (Pekarek Doehler & Ziegler 2007). Participants organize the activity of content and language through interaction, they use the L2 not only to communicate about something but also to project their situated identities (Pekarek Doehler & Ziegler 2007). The interplay between content, language,

and the organization of classroom interaction are constantly present (Pekarek Doehler & Ziegler 2007). Other researchers have reinforced this argument (Wells 2007; Morton 2012) by claiming that language use aims at achieving pedagogical goals, which are normally the learning of knowledge and skills relevant to the subject. Thus, dichotomising CLIL classes into *content* and *language* prevents us from seeing how CLIL contexts can be beneficial for both language development and content learning. When researching on CLIL, combining different approaches may be necessary to understand the integration of language and content, and the interrelated and various complex processes that occur simultaneously in a CLIL classroom (Nikula, Dalton-Puffer & Llinares 2013; Llinares, Morton & Whittaker 2012).

On the integration of language and content in CLIL, special mention needs to be made of Llinares, Morton and Whittaker's study (2012) on interaction and its role in the learning of both components in integration. They show how, in the instructional register, talk moves from everyday experiences to scientific knowledge, in the same way that students' foreign language use needs to move from everyday and context-embedded language to academic and written language. In this study, the authors also illustrate the opportunities that a dialogic-interactive mode of communication offers in order to engage CLIL students cognitively. In terms of IRF patterns, the study emphasizes and instantiates a variety of scaffolding strategies that CLIL teachers can use to support students' development of linguistic resources in the foreign language in order to express different kinds of meaning relevant for the subject. More recently, Llinares (2015) further illustrates the role of interaction in *how* to integrate content and language in CLIL (see also Nikula et al. 2016 on integration in classroom practices).

Also, taking a socially-oriented view towards learning, several studies on CLIL have explored social and interpersonal issues in interaction, such as Dalton-Puffer (2005, 2007), Dalton-Puffer and Nikula (2006), Nikula (2002, 2005). A general conclusion is that teachers should be aware of the restrictions that educational discourse impose on students, and they need to try to provide students with opportunities for engaging in different types of talk, in this way providing them also with space for engaging in FL use (Nikula, Dalton-Puffer & Llinares 2013).

In CLIL classrooms, there is a certain degree of tolerance for the construction of shared understanding (Dalton-Puffer 2009), which may explain why CLIL classes are rich in the use of scaffolding (Llinares, Morton & Whittaker 2012; Evnitskaya 2012), and teacher's monologues seem to be infrequent (Dalton-Puffer 2007). Some studies show a variety of interactive techniques used by CLIL teachers, such as passing students the responsibility for solving doubts and building bridges between unfamiliar and abstract concepts and students' everyday experiences (Llinares, Morton & Whittaker 2012; Evnitskaya 2012). CLIL teachers, as any

other teacher, need to promote students' participation in the co-construction of learning and understanding, but in addition, CLIL teachers also need to have a wide variety of interactional strategies in order to aid students' understandings of concepts which are co-constructed in a foreign language (Evnitskaya 2012).

Dialogic interactive teaching is fundamental in any type of classroom, but even more so in a CLIL classroom, due to its dual focus (Haneda & Wells 2008): students access comprehensible input and produce comprehensible output, they learn appropriate communicative strategies, and they use the language for negotiation, as different subject matter concepts and ideas are expressed and disagreements might arise. As studies comparing EFL and CLIL conclude, CLIL learners seem to be more active participants in interactions and they can use the foreign language in relevant situations, yet if teacher-centred practices predominate, these opportunities will be hindered (Nikula, Dalton-Puffer & Llinares 2013).

CLIL classrooms are interactional spaces in which L2 development can occur if participation and engagement are assured (Morton 2012). However, some studies are not so optimistic about content-based classrooms being beneficial for second language acquisition (Swain 1988; Musumeci 1996; Pica 2002). Swain (1988) claimed that good content teaching does not equal good language teaching, and she observed that in immersion contexts students were not engaged in extended discourse, correction of content predominated over correction of form, and form-meaning relationships were not made clear. All this would have to be changed in order for good content teaching to become good language teaching as well. As opposed to Swain (1988), Pica (2002) found that students in immersion classes were engaged into extended discourse. For her, the problem was that this discourse focused on meaning instead of language (Pica 2002). Swain's findings were also echoed in some studies of EAL learners (Cameron, Moon & Bygate 1996; Leung 2001; Creese 2005): EAL students in mainstream classrooms, instead of using cognitively complex language and precise vocabulary, used minimal and simple language. This shows that despite their language difficulties, their language learning needs are not paid attention to (Leung 2001; Creese 2005). In immersion research, Lyster (2007) advocated for a counterbalanced approach, meaning that focus on form needed to be included along with focus on meaning. More and more researchers advocate for this approach in CLIL settings too (see Llinares, Morton & Whittaker 2012).

3.2 ASSESSMENT FOR LEARNING (AFL)

This section is devoted to the explanation of AfL theory. First, its most important characteristics will be described. Next, some considerations about the implementation of AfL will be reflected upon, especially related to practical and political issues; some of the most important research projects developed in an AfL framework will be briefly outlined. Different types of AfL and different types of evidence accounted for in the literature will be next explained. Three of the main characteristics of AfL will be explored in more detail: feedback, self-assessment and peer assessment. Then, the relevance of different subjects areas with respect to AfL will be presented. Finally, a relationship between AfL and CLIL will try to be established, giving relevance to those aspects of AfL which are especially important in CLIL classrooms.

3.2.1 What is AfL?

Many authors have agreed that all assessments are social processes (Black & Wiliam 1998a: 56; Broadfoot & Black 2004; Arkoudis & O'Loughlin 2004; McNamara 1997). "Assessment is not an exact matter […] It is less a technical matter of measurement and more a human act of judgement" (Harlen & James 1997: 7). To make the judgement that assessment implies, certain parameters are necessary. These parameters interplay with one another and they can be implicit or explicit (Taras 2005: 2; Black & Wiliam 1998a: 8; Sadler 1998: 2). Explicit parameters yield transparency to the assessment process, while with implicit parameters the salient features for the assessor are not known.

Several terms have been used to refer to Assessment for Learning (AfL): formative assessment, classroom-based assessment, formative teacher assessment, or teacher-based assessment, among others. Scriven (1967) was the first one to talk about formative and summative evaluation in relation to the curriculum. Later on, Bloom (1969) applied these concepts to student learning (using tests with formative purposes). Black and Wiliam (1998a: 7-8) define AfL as "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged". According to Wiliam (2006: 285; see also Black & Wiliam 2009), evaluation is formative "if and only if something is contingent on their outcome, and the information is actually used to alter what would have happened in the absence of the information". That means that assessment is only formative when the evidence gathered is used to adjust teaching in order to meet students' needs (Black et al. 2003; Wiliam 2006). In other words, "[a]ny assessment for which the first priority in its

design and practice is to serve the purpose of promoting students' learning" is Assessment for Learning (Black et al. 2004: 2). Assessment activities help learning because they provide information that both teachers and students use as feedback and, as a result of it, they may modify the teaching and learning activities in which they are engaged. AfL implies, then, gathering, interpreting, and acting on the evidence to improve student learning (Bell & Cowie 2001).

Since AfL's main goal is to promote students' learning, several researchers believe it is at the heart of effective learning and teaching (Leung 2004: 23; Tunsall & Gipps 1996; Black 2003; Leahy et al. 2005; Gearhart et al. 2006). It can be defined as an assessment process (Black 2003: 2; Wiliam 2000b: 15), or as a classroom learning and teaching process (Torrance & Pryor 2001: 624, Black& Wiliam 1998: 8). In either case, the process involves both teacher and pupils. As Brockhart puts it, AfL is "as much about learning as it is about assessment" (Brookhart 2009: 1). AfL is a process, not an instrument (Heritage 2010). Teaching and learning and assessment are integral parts of the same process (Black & Wiliam 1998a, b; Leung 2004; Leung & Mohan in press; Rea-Dickins 2001).

It is compulsory to start this theoretical review with Black and Wiliam's work, especially their seminal paper in 1998, which is a review of the literature in which AfL is shown to provide learning gains. They state that the focus needs to be changed from tests to classroom interaction (Black & Wiliam 1998a: 7). They base their review on previous ones, namely those of Natriello (1987) and Crooks (1988), as well as some previous articles that had addressed the importance of AfL and its implementation. Results reported in Black and Wiliam's review clearly demonstrate that AfL upgrades student learning. In their own words, "the research reported here shows conclusively that formative assessment does improve learning. The gains in achievement appear to be […] amongst the largest ever reported for educational interventions." (Black & Wiliam 1998a: 61). Results are so overwhelmingly conclusive because the review includes studies that vary from 5 year-olds to university students, across a wide range of subjects and across different countries. Also, most of the studies in Black and Wiliam's review are not experimental (requiring a special intervention). That is, they were studies that reflected real teaching, as in the present study. In fact, Black and Wiliam (1998a: 59) urge for studies in which quantitative perspectives are combined with qualitative ones about the processes and interactions in the classroom, which is also the method applied in the present study (see also Edelenbos & Kubanek-German 2004; Ellis 1984; Hall & Walsh 2002; Cazden 1986; Mehan 1979).

AfL's main characteristics can be summarized as follows (Black & Wiliam 1998a, b; Bell & Cowie 2001; Assessment Reform Group 1999, 2002):

1. It is *central* to classroom practice.
2. It is *embedded* in teaching and learning, and it has to be included in *planning*.
3. It focuses on how students *learn*.
4. *Every student* can improve.
5. It recognizes the full range of *achievements* of all students.
6. Teachers *adapt* their teaching on the basis of the evidence they gather from students. Data are reflected upon and reviewed both by teacher and students.
7. Students are given *feedback* to improve learning.
8. There is active participation of students through *self-assessment* and *peer assessment*.
9. Teachers *share criteria* with pupils and help them recognize the *goals* they are aiming for.
10. It fosters and takes into consideration learner *motivation* at the same time that it is sensitive to any emotional impacts it may have.
11. It improves the quality of classroom *discourse and questioning*.
12. It generates *challenging and rich* tasks (see also Black et al. 2011).
13. It generates *responsiveness* to the information gathered or elicited. Teachers need to respond to and interact with students' thinking, which they have previously elicited.
14. It is a *tacit* process: many times teachers are not aware they are doing AfL (especially unplanned and interactive AfL)
15. It is recognized as a *key professional skill* for teachers.
16. It can be seen as a *sociocultural and discursive activity* in that the context (for instance, learning situations, learning activities, topic of the lesson, teachers' knowledge of students, teacher's professional knowledge and skills, teacher's purposes, etc.) has an influence on AfL (how the information is elicited, interpreted and acted upon).

Based on these characteristics a number of powerful teaching strategies in AfL have been identified:

1. *Clarifying and sharing learning intentions and criteria* with students through discussions about quality of other students' work (see also Hattie & Timperley 2007; Van Lier 1996). Several studies show how the ambiguity and misunderstanding of criteria lead to negative results. In turn, explicitness of criteria could have a negative effect, as some students might study just what is being assessed (Deutsch 1979; Natriello 1987). Barnes (1975) affirms that students will perform better if they know what is expected from them.

2. Planning and carrying out *effective classroom discussions, questions, and learning tasks*.
3. *Elicitation of the right evidence* (Wiliam 2006: 285; Black & Wiliam 2009).
4. *Interpreting the evidence correctly* (Wiliam 2006; Leung 2004; Rea-Dickins 2001). Teachers must analyse student responses in relation to what they tell about student conceptions, misconceptions, and knowledge. To do this analysis, it is essential for teachers to have domain knowledge, and hence, focus on deep understanding rather than superficial, mere recalling of facts (Crooks 1988; Harlen & James 1997). This interpretation of student responses needs to be done in a moment-to-moment basis when dealing with on-the-fly assessments. When the assessment is curriculum-embedded, it gives the teacher more time to examine students' work (it might be done at the end of a unit). If teachers interpret the evidence in the wrong way, it can lead to errors in what the next instructional steps should be. The interpretation of this evidence (student responses) is what is later realized through feedback.
5. *Provide feedback to move learners forward*, so that they know what to do next and how to do it (see also Black 2001; Black, Harrison, Lee, Marshall & Wiliam 2003).
6. *Matching the instruction to students' gaps*. It cannot be too close because students will be bored or too far because students can be frustrated (Sadler 1989). It must entail the appropriate demands for learners (see also Leung 2004; Rea-Dickins 2001; Llinares et al. 2012; Harlen & James 1997; Torrance & Pryor 2001; Leahy et al. 2005; Lynch 2001; Hattie & Timperley 2007). Ideally, new learning will be internalized by pupils and thus become part of their development.
7. *Promoting self-assessment*, students as owners of their learning.
8. *Activating students as instructional resources for one another* (peer assessment) (Sadler 1989; Leahy et al. 2005; Harlen & James 1997). Teaching students how to assess their own work and their classmates' (identify what they did wrong or did not understand, know how to improve, etc.) can be done through questions for opinions and meta-cognitive questions (do you think that…? Why do you think that…?) (Heritage 2007). Teachers must teach student self-assessment techniques in a collaborative way, developing a shared understanding of students' current state of learning and what they can do to improve it (Sadler 1989; Heritage 2007).

The positive effects described in Black and Wiliam's review were sustained over time, and were even observed when student achievement was measured in external standardized tests (Leahy et al. 2005; Wiliam, Lee, Harrison & Black 2004; Black et al. 2004; Shepard 2001; Brookhart 2004; Black & Wiliam 2006). No reports specifying negative effects of AfL have been found (Black & Wiliam 1998a: 17), although some key weaknesses of teachers' practices have been

identified (Crooks 1988; Black 1993b; Black & Wiliam 1998a: 17-18). For example, in weak practices teachers promote rote learning; assessment questions are not used, not reflected upon, and not discussed with students; grading is emphasized over learning; competition is encouraged among learners instead of personal and individual improvement. Both in questioning and written work, teachers tend to aim at low-level goals (Stiggins et al. 1989; Black & Wiliam 1998a: 18), and so students do not risk engaging in demanding cognitive activities (Duschl & Gitomer 1997).

Even though educational gains when implementing AfL seem apparent, summative assessment is also necessary, especially to monitor overall levels of achievement (Leahy et al. 2005; Heritage 2010; Black et al. 2004; Taras 2005; Harrison & Howard 2009; Black et al. 2011). In fact, Crooks (1988) advocates for a balance between formal and informal evaluation. Nonetheless, it is widely known that testing can produce washback effects, that is, negative effects for instruction, as teachers teach to the test, and so students' motivation and time available for AfL decrease (Llinares et al. 2012; Harlen 2005; Black et al. 2003; Broadfoot et al. 1998; Black et al. 2011; Assessment Reform Group 2002; Harlen & Deakin Crick 2002). In addition, the information that a test provides is hard to use because of two main reasons (Leahy et al. 2005): a) that information comes delayed, that means, by the time we get the test results instruction often cannot be modified; b) the information cannot be used to inform about specific weaknesses, as each skill or goal covered is only allotted a small amount of testing time.

A balance between AfL and summative assessment may be difficult. A way for the combination of both is to use summative assessment for formative purposes. For tests to serve a formative assessment function, they should not just give information about a final grade, but they should also provide feedback about learning (Black & Wiliam 1998a: 36; Black & Wiliam 1998b: 144; see Gearhart et al. 2006, Black et al. 2003, and Dempster 1992 as examples of teachers using summative tests for formative purposes). Both Maxwell (2004) and Black et al. (2003) offer examples of formative use of summative assessments and argue that it is one of the most successful techniques, along with questioning, feedback and self-/peer-assessment, reported by teachers when implementing AfL.

As a summary of this section,

> much of what teachers and learners do in classrooms can be described as assessment. That is, tasks and questions prompt learners to demonstrate their knowledge, understanding and skills. What learners say and do is then observed and interpreted, and judgements are made about how learning can be improved. These assessment processes

> are an essential part of everyday classroom practice and involve both teachers and learners in reflection, dialogue and decision making (Assessment Reform Group 2002).

Once AfL has been defined and its benefits have been highlighted, the next subsection will deal with some practical considerations that are needed when implementing AfL and some political considerations that hinder AfL implementation.

3.2.2 Practical and political considerations on the implementation of AfL

Broadfoot and Black (2004: 19) argue that we live in an "assessment society", in which there is a strong "belief in the power of numbers, grades, targets and league tables to deliver quality and accountability, equality and defensibility". There is enough information and evidence to provide guidance to practical AfL action, yet the actual level of implementation is low (Black 1993b; McCallum et al. 1995; Wiliam et al. 2004; Tunsall & Gipps 1996; Torrance & Pryor 1998). In addition, there are social and psychological barriers to AfL (Dwyer 1998). For AfL to be implemented, big changes need to be made in teacher training programmes in relation to pedagogy and teachers' perception of their role in relation to students (Perrenoud 1991; Leahy et al. 2005; Heritage 2007, 2010; Black et al. 2004).

For AfL implementation, teachers need to be involved, as well as researchers and administrators (Black & Wiliam 1998a, b; Leung 2004; Sebatane 1998; Dwyer 1998; Broadfoot & Black 2004). Teachers need examples of practical implementation and an opportunity to share ideas and experiences with each other, creating professional communities of practice (Black & Wiliam 1998a: 146; Leung 2004), as well as they need support and training (Black & Wiliam 1998b; Black et al. 2011). As some authors have argued, AfL takes more time and requires greater skills from teachers. For these reasons, it is logical that changes will happen slowly, as deep-rooted habits are difficult to change, and each teacher should incorporate these changes into their classroom practice in their own way (Black & Wiliam 1998b; Broadfoot et al. 1998; Black et al. 2004; Harrison & Howard 2009). However, programmes which are very different from current practices or which involve extra work will very likely not be implemented (Guskey 1985). AfL changes will take more time in class, so teachers need to believe that the investment is worth for the future learning of students (Black & Wiliam 1998b; see Black & Harrison 2001b for reports of teachers satisfied with AfL practices), and not think that AfL is something taking time away from them and imposed on them from the outside (Heritage 2007; Hall et al. 1997; Black et al. 2011). Black et al. (2004) recommend that teachers start including AfL techniques at the beginning of the year (because, otherwise, it is very difficult to change habits that have been going on for some months) and in one group (or in one subject) to be able to

compare the results with other groups or subjects. Changes will be difficult for students as well, as they may be afraid of them until they experience their positive effects (Black & Wiliam 1998b: 144).

Renegotiating the learning contract is another requisite to implement AfL (Perrenoud 1991). This means that teachers also have to believe that teaching and learning is not just a matter of transmission (Black & Wiliam 1998b; Black et al. 2004; Barnes 1975; Hattie & Timperley 2007; Alexander 2004; Van Lier 1996; Barnes 1975). In other words, AfL is jointly constructed by all participants in the classroom, which makes classrooms no longer teacher-centred but student-centred (Heritage 2010). As Easley and Zwoyer (1975: 25) state (in Crooks 1988: 469):

> If you can both listen to children and accept their answers not as things to just be judged right or wrong but as pieces of information which may reveal what the child is thinking you will have taken a giant step toward becoming a master teacher rather than merely a disseminator of information.

There are important difficulties in the implementation of AfL. For example, in the UK, although AfL is recognized, governments commit to high-stakes external testing, and thus, national and local requirements for certification and accountability affect greatly and negatively AfL practice (Black & Wiliam 1998a, b; Rea-Dickins 2001; Rawlins 2010; Black 2012). As in the UK, governments in different countries are more concerned about high-stakes assessment systems that account for individual attainment than about AfL (Leung & Rea-Dickins 2007; Brindley 1998). In 2004, teacher assessment was reported and taken into account again, after many years of darkness and unrecognition and after failure in standardized assessment to meet the official targets (teachers were teaching to the test and narrowing the curriculum, and they were blamed for the low achievement of pupils in literacy) (Leung & Rea-Dickins 2007). However, according to Heritage (2007), teachers still think assessment is something external to their everyday practice. The reciprocal relationship between instruction and assessment has been lost from sight, and assessment is viewed as something which is in competition with teaching (see also Rawlins 2010; Harlen & Deakin Crick 2003, Wiliam, Lee, Harrison & Black 2004). An additional problem, Heritage claims (2007: 1-2), is that teachers learn how to teach, but not how to assess, since assessment is not part of preservice and in-service courses (at least in the USA) (see also Natriello 1987). In sum, assessment, curriculum and instruction must be aligned in a coherent way in national systems (Black 2012), and this should be advocated for in every country.

There are other factors intervening in the relegation of AfL, such as large size classes and heavy workloads (Carless 2007: 173). Particularly in secondary education in the UK, AfL is not well

understood by teachers and hence not widespread (Rawlins 2010; Black & Wiliam 1998a, b; Crooks 2006; Wiliam, Lee, Harrison & Black 2004). One way to solve this problem is to get teachers to participate in both summative and formative assessment (Black & Wiliam 1998a; Brindley 2001; Broadfoot & Black 2004; Black et al. 2011). Researchers in UK claim that teachers should be more involved in the summative assessments that take place for accountability reasons, since no one better than them know children's progress, and it would be one way of reconciling tensions between formative and summative assessment, as well as a way of reducing washback effects (Black & Wiliam 1998b; Broadfoot & Black 2004; Brindley 2001; Black et al. 2011; Black 2012).

Although standards are trying to be raised in many different countries, what drives learning is what teachers and students do in classrooms, inside the black box. Therefore, standards have to be raised in classrooms, not in formal testing (Black & Wiliam 1998b: 140; Clarke 1998; Harlen & Winter 2004). Having standards allows assessment to be aligned with instruction (McKay 2006). They also provide a common and shared reference framework not only for teachers but also for administrators and assessors, facilitating communication about learning and progress (McKay 2006). Another advantage of standards is that they offer an objective basis upon which administrators' decisions about a programme (needs, resources, etc.) are based (McKay 2006). Extensive research shows that higher standards lead to better student performance, although the positive results are not true for all contexts and students (Natriello 1987). If learning is improved, performance in high stakes tests and examinations will be higher as well (Ruiz-Primo & Furtak 2007). However, links between AfL practices and measures of student learning are only beginning to be put into practice (Black & Wiliam 1998a; Ruiz-Primo & Furtak 2007).

In spite of the barriers found in the implementation of AfL, Black and Wiliam's 1998 review led to some political changes in the UK, and it also supported other work on formative assessment in other countries (Broadfoot & Black 2004: 10; Black & Wiliam 2003). Some reforms of teacher assessment have been undertaken in France (Broadfoot 1996) and in French Canada (Dassa 1990). This movement towards AfL has been translated into over 40 papers on how AfL is implemented in the classroom (Broadfoot & Black 2004). Despite this shift in attention to AfL, however, it seems that the interest in AfL has not been so much reflected in second and foreign language contexts (Rea-Dickins 2008) nor in CLIL classrooms. Likewise, Spain is not among the countries that overtly promotes and supports AfL.

As Black and Wiliam (1998b: 148) warned, the success in the implementation of AfL is only possible if government, administrators, researchers, and practitioners work collaboratively. Several years ago, important researchers, especially in the UK, already started different research projects in order to help schools and teachers implement AfL. These will be briefly reported in the next subsection.

3.2.3 Paramount AfL research projects

Within this subsection, the most important research projects on AfL will be explained. The theory of AfL was born in the UK, and so was its research. The first AfL research project was carried out by researchers from King's College London. The second one was also developed in the UK, under the coordination of Rea-Dickins and Gardner. It is an *Early Years Intervention Project* and it is specifically designed to support learners for whom English is an Additional Language (EAL) (Rea-Dickins & Gardner 2000).

With the ultimate objective of enhancing the implementation of AfL, a group of researchers at King's College London started an AfL project in 1998 and have published widely on the results (Black & Wiliam 2003; Black et al. 2003; Wiliam et al. 2004; Harrison & Howard 2009; Hodgen & Harrison 2005; Black & Wiliam 2009; Black & Harrison 2001a, b). This group of researchers worked with 12 teachers of Science, 12 of Mathematics, and 12 of English (incorporated at a later stage) to implement AfL in 6 different schools (Black & Wiliam 2003; Wiliam et al. 2004). These teachers were given support on AfL for a six-month period, and the next year, at the beginning of the school year, they started practising the different techniques of AfL. There was always a control group, sometimes it was a parallel group taught by another teacher, other times an equivalent class taught in the previous year by the same teacher and other times a parallel or another class taught by the same teacher. Overall, the positive mean effect size at the end of the first year of intervention was 0.32 (see also Black & Wiliam 2003). After two years of intervention, results in terms of AfL effects and gains in test performances were very positive (Harrison & Howard 2009). They could not affirm that it was for sure due to the implementation of AfL, but that was the most likely cause (Black & Wiliam 2003). Changes in teachers' assessment practices were slow and they were noticed at the end of the year, even though the implementation started at the beginning of the year. The intervention in this project was translated in the teachers using the AfL techniques in all their classrooms and it changed the view they held of themselves as teachers (Wiliam et al. 2004: 14; Black et al. 2003; Black & Harrison 2001a). Successful AfL practices were shown over a range of different subjects and ages (Harrison & Howard 2009). Results took 2 years to be achieved, but they were positive.

A second important project was the one developed by Rea-Dickins and Gardner (Rea-Dickins & Gardner 2000; Gardner & Rea-Dickins 2001; Rea-Dickins 2001), who investigated AfL in elementary classrooms, exploring the assessment process and the quality of teacher assessment, as well as classroom interaction. Rea-Dickins (2001) investigated into the assessment process and the assessment cycle (identifying different stages) in English as an Additional Language (EAL) learners in mainstream classrooms. Apart from the regular teacher, these classrooms also have language supporters. She drew on Hall et al. (1997), who propose different linear stages for teacher assessment, O'Malley and Valdez Pierce (1996), who identified six purposes of assessment with EAL learners, and Clarke (1998), who worked with primary teachers and organised assessment according to its function, proposing different strategies for different stages. Especially relevant for Rea-Dickins' model was Hall et al.'s (1997) developmental model of assessment, since it was the basis for her proposal. Hall et al.'s model comprises the following stages: *assessment planning stage*, before the year begins; *observation stage*, at the beginning of the year, when the teacher gets to know her/his class and the learning states of their pupils, with a potential formative purpose; *specific task stage*, when work and individual needs are met, there is a strong curriculum focus, and assessment centres on progress in relation to curricular criteria; *continuous review stage*, which is the longest because it goes on throughout the whole year, with the teacher recording and gathering and judging evidence; finally, the *levelling stage*, which happens after a unit and consists of allocating a level of attainment to each learner. In Rea-Dickins' (2001: 435) classification, stages are not necessarily linear: planning, implementation, monitoring and recording, and dissemination (see also McKay 2006 for different stages in the assessment cycle). All the stages are interconnected, they are iterative and more subtle in AfL than in testing (McKay 2006). Similarly, she identifies three types of assessments that could be placed on a formal-informal continuum (Rea-Dickins 2001), the first one being more formal and similar to formal language tests, the other two more informal and focusing on interaction and teacher intervention during it. In informal AfL, we find scaffolding and teachers' interventions being partly planned and partly responsive to the interaction (Rea-Dickins 2001; see also Bell & Cowie 2001). All in all, Rea-Dickins (2001) analyses interaction in relation to the four stages of assessment and links them to the three types of assessments she identifies.

Thanks to the proliferous research on AfL, especially in the UK, different types of AfL and different types of evidence have been identified, as the following subsection will describe.

3.2.4 Types of AfL and types of evidence

Several authors have said that AfL can be seen as a continuum where we have formal and informal at the two ends (Bell & Cowie 2001; Shavelson et al. 2003; Rea-Dickins 2001). Formal AfL is planned beforehand. It might consist of activities, questions, quizzes, etc. that typically focus on one aspect of learning, so the information is very specific. Formal AfL enables teachers to check understanding and plan on the next teaching steps that will move learners forward (Ruiz-Primo & Furtak 2007; Furtak & Ruiz-Primo 2008). It enables teachers to step back, and analyse and interpret the information collected, and based on it, plan the appropriate action (Ruiz-Primo & Furtak 2007). On the other hand, informal AfL is more spontaneous. It can take place during any teacher-student interaction and it is "embedded and strongly linked to learning and teaching activities" (Bell & Cowie 2001: 86; see also Duschl 2003; Shavelson, Black, Wiliam & Coffey 2003). This type of AfL is the contingent scaffolding or support given by the teacher at the moment to the students for them to complete learning tasks or show their understandings (Sharpe 2006; Tharpe & Gallimore 1988; Wood et al. 1976; van de Pol et al. 2010; Gibbons 2003). The information gathered by this type is momentary, and teachers have to interpret and act on it immediately without interrupting the flow of the classroom discourse. It allows teachers to complete several feedback cycles in a short period of time (Ruiz-Primo & Furtak 2006; Furtak & Ruiz-Primo 2008). This action is flexible, for example, when suddenly a student's misunderstanding is expressed, the teacher can respond with another question, eliciting other classmates' views, repeating an activity, etc. (Ruiz-Primo & Furtak 2007). Informal formative assessment gathers information continuously about the learning of students in any teacher-student interactions that go on in normal classrooms. These types of interactions are also called assessment conversations (Duschl & Gitomer 1997; Bell & Cowie 2001).

Intimately linked with the previous classification, Cowie and Bell (1999) made the distinction between planned and interactive AfL. Interactive AfL takes place during interaction and the information is ephemeral. That is why teachers have to make quick decisions, making use of the context and their knowledge of students. It serves to refine teachers' short-term goals. Noticing is the first part of the process, followed by recognising (often triggered by incorrect or unexpected answers), for which teachers have to use prior knowledge about the student. Recognising also implies interpretation and appreciation of implications (Cowie & Bell 1999). To do that, teachers have to make qualitative judgements, for which they have to have a concept of quality appropriate to the task and the ability to judge students' work (Sadler 1989). These judgements invoke fuzzy criteria, more context-dependent than predetermined (Sadler 1989).

As Mehan states (1979: 112),

> before the lesson the teacher may have the criteria she intends to use to judge student responses, however, these may change during the lesson, influenced by events which emerge within the situation; who is answering, when in the lesson, what has gone before, what the child has done before.

The third and last part of interactive AfL is responding to what is noticed and recognized. In the present study, the focus will be on informal and interactive AfL.

Another classification is the one developed by McKay (2006), who distinguished two types of AfL: a) on-the-run, with formative purposes, where observation is paramount and unexpected, and based on scaffolding; and b) planned, with either for formative or summative purposes, it involves thought beforehand to collect the information required, and for this to be valid and useful. Questioning, seeking clarification, pushing learners forward in their learning are some types of interventions in on-the-run assessment (Rea-Dickins 2001: 437-438). On-the-run assessment is equivalent to interactive AfL (Cowie & Bell 1999) and informal AfL (Rea-Dickins 2001), which represent the kind of AfL studied in the present study.

In a similar line, Torrance and Pryor (1995) distinguished two approaches within AfL: behaviourist (highlighting measurement against objectives) and social constructivist (integrating assessment as part of the learning process). Torrance and Pryor (2001) distinguished between convergent and divergent assessment. In convergent assessment, which is carefully planned, what is important is to find out whether the student knows, understands or can do a given thing. It is normally achieved through closed questions and tasks. On the contrary, in divergent assessment the important thing is to discover what the student knows, understands or can do. It is less planned, and is achieved through open questions and tasks. It involves a constructivist view of learning (as opposed to a behaviourist one, involved in convergent assessment), teaching in the ZPD (Vygotsky 1978), that means, scaffolding students in order to extend their current abilities and thus expand their learning. In this way, assessment is jointly constructed and accomplished by the teacher and students through interaction. This kind of assessment is in line with informal and interactive AfL and, hence, it is the type of AfL investigated in this research.

A different but related classification of AfL types is Heritage's (2007). She differentiates three types of AfL, depending on the strategies used to gather evidence of learning:

1. *On-the-fly assessment*: it takes place spontaneously during classroom interaction (equivalent to informal, interactive, on-the-run AfL)
2. *Plan-for-interaction*: teachers plan beforehand what questions they are going to ask and how they are going to elicit students' thinking.
3. *Curriculum-embedded assessment*: further subclassified into those that "teachers and curriculum developers embed in the ongoing curriculum to solicit feedback at key points in a learning sequence and those that are part of ongoing classroom activities" (Heritage 2007: 2).

Wiliam (2006) differentiates AfL into different types depending on time: short-cycle AfL (seconds or minutes), long-cycle AfL (years, semesters), and medium-cycle (teaching units). He claims that most of the studies have focused on medium-cycle AfL, even though this type has been proven to be the one that yields less positive results in relation to student learning (Black & Wiliam 1998a), maybe because they were summative in their origin (final grade at the end of a semester or at the end of the year) or because of the pressure to finish the syllabus (Wiliam 2006: 288). Other studies which have focused on short-cycle AfL have shown more impact on student learning (e.g. Wiliam, et al. 2004; Leahy et al. 2005). Only when long- and medium-cycle AfL require changes in actual classroom practice there seems to be a positive impact on student learning (Wiliam 2006). The focus of the present study will be both short-cycle and medium-cycle, as we deal with individual lessons of 55 minutes as well as with complete teaching units.

Another study that identifies different types of AfL is Ruiz-Primo, Shavelson, Hamilton and Klien (2002), who pinpointed five levels of assessment, each of them indicating a greater distance from a certain curricular activity and forming a continuum: immediate, close, proximal, distal, and remote. The various levels were established according to their distance to the curriculum. Immediate assessments are students' products from the enactment of the curriculum; close ones parallel the content and activities of the unit; distal assessments tap the knowledge and skills which are relevant to the curriculum; and remote assessments reflect national standards in a certain knowledge domain. Their results show (Ruiz-Primo et al. 2002) that instruction impacted on students' performance, since significant differences were found between the pre-test and post-test at the close, proximal, and distal levels of assessment. Proximal assessments did not show as much impact of instruction as close ones. The effect of instruction decreased as the level of the assessment increased. However, the correlation between

performance in the post-test and the immediate level of assessment was weak. Immediate-level assessment is what concerns the present research: "collective participation in discourse practices […] as a particular lesson is enacted" (Hickey et al. 2006: 188). Teachers and students can directly observe discourse, and formative feedback can directly enhance it (Hickey et al. 2006: 9). Its timescale is of minutes, it is organized on a moment-to-moment basis.

As already explained, when implementing different AfL types, teachers use evidence coming from students in order to assess the learning gap and the best intervention to close it. This evidence can be of different kinds as well. For instance, Wiliam and Black (1996) differentiate between purposive and incidental evidence. The former is the result of a specific action on the part of the teacher (normally) to provide evidence about students' knowledge in a particular area. This is normally done through direct questioning. However, disclosure of the assessment has to be taken into consideration . This is the extent to which an assessment can be "relied upon to yield evidence of attainment where it exists" (Wiliam & Black 1996: 541). For example, a concrete question might not receive any answer, but a slight change in the question may elicit the evidence that the teacher is looking for. Incidental evidence is that which is generated spontaneously, without the teacher specifically asking students to do something or to answer some specific question. It is a more robust type of evidence than purposive one, since the student spontaneously decides to use a specific skill and does so successfully, which is probably evidence of deeper learning.

Wiliam and Black (1996) also differentiate the form in which evidence is elicited. In this way, evidence can be permanent (evidence relied on in summative assessments) or ephemeral (essential for AfL) (see also Bell & Cowie 1999, 2001; Duschl 2003; Shavelson, Black, Wiliam & Coffey, in press; Ruiz-Primo & Furtak 2006, 2007; Furtak & Ruiz-Primo 2008).

The interpretation of the evidence is done by the teacher, and can be done in different ways. The teacher, through interpretation, determines whether there is a gap or not (Wiliam & Black 1996: 4). The teacher gathers and interprets evidence of attainment based on a model: if students' answers fit the model, they'll be regarded as having understood the topic. Interpretations are means to an end rather than an end in itself. In other words, if assessment does not lead to changes for improvement, there is no point in making an assessment (Wiliam & Black 1996: 6).

The information gathered through AfL comes from observing, questioning, listening to students' contributions, reviewing their work and using students' self-assessment (Harlen & James 1997). This information may be used immediately or later to plan learning opportunities (Harlen 2005), but the ultimate user of the assessment information should be the learner (Harrison & Howard 2009).

In AfL, then, evidence is crucial. Without evidence, teachers would not have any means on which to base their assessments of learning gaps, and therefore, they would not have any foundation to justify their teaching interventions aimed at improving student learning. One of the possible interventions teachers have at their disposal to help students close learning gaps is providing them with feedback, which will be the focus of the next subsection.

3.2.5 Feedback and AfL

To begin with, it is important to distinguish between the concept of feedback dealt with in section 1.1.2.2 of this chapter and the concept of feedback presented in this section. The concept of feedback explained above is primarily concerned with interaction. Also, it usually provides students with information that a particular utterance is incorrect, normally in terms of second/foreign language learning, as the concept was born in SLA research. On the other hand, the concept of feedback in AfL is more general, in the sense that it is not necessarily limited to interaction and it offers information about how to improve the learning process, whether it is content learning, language learning, or any other type of learning.

Feedback, as defined within the AfL approach, refers to the "information about how successfully something has been or is being done", it is the decisive element to assist learning (Sadler 1989: 120). According to Ramaprasad (1983), feedback is the information about the gap between the actual level and the reference level. Teachers use it for diagnosis and remediation, and students for monitoring their weaknesses and strengths (Sadler 1989). Ramaprasad (1983) emphasizes that feedback is only feedback and formative when it is used to alter the gap, and not if it is passed to a person who does not have the power or knowledge to change the outcome (see also Sadler 1989: 121; Wiliam & Black 1996; Rea-Dickins 2001; Hattie & Timperley 2007). The best feedback is given when the gap is clearly identified (Weeden et al. 2002: 109; Hattie & Timperley 2007). Sadler (1989: 143) identifies three conditions for effective feedback: "[feedback] requires knowledge of the standard or goal, skills in making multi-criterion comparisons, and the development of ways and means for reducing the discrepancy between what is produced and what is aimed for". This echoes Ramaprasad (1983), who says that for feedback to take place, there must be data on the reference level, data on the actual level, and a mechanism to compare both and in this way establish the gap. To assess the gap, differences are made between the nature of the gap (qualitative) and the magnitude of the gap (quantitative) (Ramaprasad 1983: 4). As Ramaprasad (1983) explains, the feedback that widens the gap between actual and reference levels is called positive feedback. On the other hand, the feedback which reduces the gap is called negative feedback (see also Black & Wiliam 1998a). This contrasts with common beliefs, more to do with the emotional connotation to the recipient of the

feedback: positive feedback would make the recipient happy; negative feedback would make him/her unhappy. Positive and negative feedback are also confused with positive and negative reinforcements (Ramaprasad 1983: 7). Both positive and negative reinforcements are negative feedback, as they try to minimize the gap between actual and reference level.

Black and Wiliam (1998a: 53) refer to feedback in the following sense:

> any information that is provided to the performer of any action about that performance […] This need not necessarily be from an external source […] nor need there necessarily be some reference standard against which the performance is measured, let al.one some method of comparing the two.

Feedback should be about how students can improve their work (Black & Wiliam 1998a, b), and therefore, their performance can be evaluated in its own terms or compared with a standard. This comparison can be made in terms of equality, distance, or diagnosis (what can I do to get there?). In this latter case, feedback has a formative function. So, as Sadler argued (1989), assessment is formative when comparison between actual and reference levels gives some information which is used to alter the gap.

The quality of feedback for guiding and improving is a key feature for AfL, since formative feedback aims at moving learning forward (Black & Wiliam 1998a; Sadler 1989; Davison & Leung 2009; Leung 2007; Harrison & Howard 2009). Feedback must be delivered to learners in terms they understand, as their knowledge of the subject is partial, and sometimes they do not know what to do with that information (Sadler 1998: 81). Sadler (1998: 82) claims that feedback should be accurate, comprehensive, and appropriate; delivered to students in an accessible way; able to inspire confidence; and able to coach and improve learning. Equally important is to allocate some time in every lesson for teachers and students to talk about the improvements needed (Harrison & Howard 2009). Students have to understand that their work is "under construction" and it is not a final product (Harrison & Howard 2009: 18). Students are gradually closing the gap, and teachers need to monitor learning as it is taking place (Harrison & Howard 2009; see also Heritage's 2007 on-the-fly assessment and Cowie & Bell's 1999 informal or interactive AfL).

Regarding the effects of teacher feedback on learning, many studies have shown that it leads to increased performance, although it also seems to depend on the feedback type. Kluger and DeNisi (1996) carried out one of the most important reviews on the effectiveness of feedback, and their results were contradictory: feedback focused on the self had negative effects, whereas feedback on the task yielded positive effects. Even giving praise or extrinsic rewards can be

negative (Ames 1992), especially if it is not related to objective feedback about the work (Black & Wiliam 1998a: 23).

Following with feedback types and their effects on learning, Butler (1987) examined four types of feedback: no feedback, grades, comments, and praise. It turned out that the comments group scored higher than the other three in measures of task-involvement. Students who were given grades and comments, ignored the comments (Butler 1988; see also Black et al. 2004; Black & Wiliam 1998a; Hickey et al. 2006; Taras 2003; Tanner & Jones 2003; Crooks 1988), so students are more productively engaged when they are given only "comment" feedback and not grades at all (Black et al. 2003; Butler 1988; Black et al. 2004; Black & Wiliam 1998a; Taras 2003; Dweck 2000; Harrison & Howard 2009; Black & Harrison 2001 as an empirical example). In the same line, Crooks (1988) found that grades and evaluative feedback did not have an impact on subsequent performance. Teachers might be afraid of not giving grades at first because it is against school policy, because of the extra time it involves and because of how students might interpret it, but this practice has yielded positive results (Black & Harrison 2001a; Black et al. 2003; Butler 1988; Dweck 2000; Harrison & Howard 2009). Unfortunately, teachers seem to think that giving detailed feedback to students' written work is a waste of time, since they invest a lot of time and students do not bother reading it (Leahy et al. 2005).

Special attention deserve Hattie and Timerley's (2007) review and Tunsall and Gipps' (1996) study on feedback types and their effects. Hattie and Timperley (2007) revised 196 studies and found that feedback had an average effect size of 0.79. This effect turned out to be greater than other variables, such as student prior cognitive ability and socioeconomic background (Hattie & Timperley 2007: 83). Hattie and Timperley (2007) distinguished four types of feedback:

1. *Feedback about the task.* It focuses on how well a task is being performed (distinguishing correct from incorrect answers, building surface knowledge). This type of feedback is also called corrective feedback and it is more effective if it is simple and if the task is simple. Likewise, it is more effective when it deals with faulty interpretations (Hattie & Timperley 2007). For this type of feedback to be effective, the climate of the classroom must be receptive to it (Hattie & Timperley 2007). Also, in this type of feedback, providing written comments is more effective than giving grades (Black & Wiliam 1998a; Crooks 1988).
2. *Feedback about the processing of the task*. It addresses the processes underlying the task. The main type is feedback related to student strategies for error detection. To enhance deeper learning, this type of feedback is more effective than feedback about the task (e.g. Balzer et al. 1989; Early et al. 1990).

3. *Feedback about self-regulation*. It deals with the way students "monitor, direct and regulate actions toward the learning goal" (Hattie & Timperley 2007: 93). It involves control, commitment, autonomy, discipline, confidence, self-direction, and self-assessment (made up of self-appraisal and self-management). Both this type and feedback about the processing of the task are essential to the implementation of AfL, since two main foci of AfL are to improve students' learning (learning seen as a process and not as a product) and to promote learners' autonomy and responsibility for their learning.

4. *Feedback about the self as a person*. Positive (praise) or negative evaluations about students are only effective when they achieve a change in student effort, attitude, or engagement. This type of feedback is widely used but it is not so effective (Kluger & DeNisi 1998; Wilkinson 1981). In fact, praise can even be counterproductive and have negative effects on students' self-assessments about their ability (Meyer et al. 1979; Meyer 1982).

Negative effects can appear when students cannot relate feedback and the cause of their poor performance (Hattie & Timperley 2007). On the other hand, feedback that relates performance to effort or ability is quite effective (Craven, Marsh & Debus 1991; Dohrn & Bryan 1994). In brief, Hattie and Timperley (2007) conclude that immediate feedback is more powerful when the feedback is about the task, and delayed feedback is more effective when the feedback is about the processing of the task. It is difficult to prove the frequency of feedback in classrooms but Bond et al. (2000) claim it is low. Often, when it is given, it is related to the task (corrective) or to the self (Hattie & Timperley 2007), whereas feedback related to the processing of the task is scarce and usually negative (Blote 1995).

The second study, which is especially relevant for AfL, is Tunsall and Gipps' (1996). They explored the types of feedback given to students (6-7 years old) in different subjects (involving 8 teachers and 6 schools) and how they understood it. Their data were classroom observations, recordings, and teacher and student interviews. They developed the following framework: a) *Feedback and socialization*, related to values, attitudes, and classroom procedures; and b) *Feedback and assessment*, in which there are four different types of feedback placed within a continuum which goes from evaluative to descriptive. Evaluative types of feedback are more affective and oriented to performance, whereas descriptive emphasize the cognitive aspect and are clearly associated to AfL (Tunsall & Gipps 1996). A combination of evaluative and descriptive types would create the most powerful context for learning (Tunsall & Gipps 1996).

Both evaluative and descriptive feedback have subcategories. Evaluative feedback is divided into rewarding/punishing and approving/disapproving:

- *Rewarding/punishing*: rewarding feedback motivates extrinsic motivation through the giving of some type of reward (stickers, stamps, etc.), and reinforces students' positive behaviours. On the opposite side, punishing is given when teachers completely disapprove students' behaviour, and is normally translated into sending students out of the class or moving them to another place.
- *Approving/disapproving*: approving feedback is when teachers positively evaluate students' work or engagement. On the negative side, disapproving feedback is negative evaluation, normally caused by bad behaviour or bad work as a result of the lack of effort or concentration.

Descriptive feedback is also divided into two subtypes:

- *Specifying attainment/specifying improvement*: when teachers use the former type, they identify successful components of attainment. It normally includes a model (standards) that specifies the basis for success. Specifying improvement is the type of feedback that describes how something that is being learnt can be improved. It therefore tends to focus on mistakes. It may also involve engaging students into self-assessment.
- *Constructing achievement/constructing the way forward*: constructing achievement is differentiated from specifying attainment in that extensive articulation of students' achievement and competence is articulated through discussions with students. It makes students reflect and emphasizes students' key role in their learning process. As for constructing the way forward, it is also provided in the form of discussion with students, and it articulates future possibilities in learning. This means that students' voice is also heard as to what they can do to improve, thus transferring them more responsibilities.

The results of this study showed that the most frequent learning opportunities came when feedback was public in whole class discussions and when the responsibility changed from teachers to students (Tunsall & Gipps 1996). All in all, as Tunsall and Gipps (1996: 400) explain, "this typology gives us an insight into the role of teacher feedback in the classroom, its role in teaching, learning and formative assessment, as well as contributing to the discussion of achievement goals".

Finally, it is interesting to investigate teachers' and students' perceptions about AfL practices. In a study carried out in mathematics classrooms in New Zealand, Rawlins (2010) concluded that whereas teachers preferred whole-class oral feedback because they thought it was effective

and students did not read written feedback, students reported that they did read written feedback. In fact, they would rather have written feedback and engage with it in peer interactions. Written feedback is considered a more concrete way of eliciting students' ideas and some research have shown that students' motivation and performance can increase when teachers provide written feedback on a task (Butler 1987; Butler & Nisan 1986). However, this is time consuming for teachers, at the same time that if they provide this written feedback in a delayed manner, it may not be instructionally relevant any longer (Furtak & Ruiz-Primo 2008). On the other hand, oral feedback is immediate and the information provided in it can be used immediately by students (Hattie & Timperley 2007; Ruiz-Primo & Furtak 2006, 2007). This disparity between students' and teachers' perceptions must be taken into account when investigating the effects of feedback in AfL.

This section has provided enough evidence regarding the importance of teacher feedback for student learning. The goal in education should be, nonetheless, to move from feedback, in which the source of information is external to the learner, to self-monitoring, in which it is the learner him/herself who generates the information (Sadler 1989: 122). Self-assessment, together with peer-assessment, then, are also crucial elements of AfL, and they will be the object of the next subsection.

3.2.6 Self-assessment and peer-assessment

Self-assessment, that is, students assessing themselves, is essential for learning and is key in AfL because learners are expected to be active agents in the assessment process, as they are also responsible for their own learning (Black & Wiliam 1998a, b; Rea-Dickins 2001, 2006; Heritage 2007, 2010; Taras 2008; Harlen 2005; Bell & Cowie 2001; Davison & Leung 2009; Harrison & Howard 2009; Leung 2004). Notwithstanding, common knowledge places the teacher as the agent of the assessment (who decides on what to assess, how to assess, etc.). And those decisions are not contested (Taras 2008). In the same line, students are reluctant to be involved in the assessment process as they consider it to be the teachers' job (Taras 2003, 2008; Cowan 2006). Both teachers and students, then, accept the state of affairs, which places teachers in a position of power over students (Taras 2008: 3; Black & Wiliam 1998a: 20). Power relations, however, can be changed if self-assessment practices are implemented, especially when they are taken into account for formal grading purposes (Taras 2008). Taras (2008) claims that the standard model of self-assessment makes students and teachers believe that students are really participating in the assessment process. However, this is not truly the case, as it does not provide access to the summative assessment process. That is why Taras' model of student self-

assessment (2008) intends to include students in the summative assessment process (hence using and integrating formative and summative assessment).

For self-assessment to work, there are several key factors. First, students need to be trained in it in order to recognize the desired and current level and what can be done to close the gap (Black & Wiliam 1998b: 143; see also Harrison & Howard 2009), as it may be hard for them to do self-assessment and they may have problems with identifying and correcting errors (Olina & Sulliven 2004; Rawlins 2010). However, even very young learners seem capable of self-assessment (Harrison & Howard 2009). Higher-order questions and students questioning themselves can serve to initiate students in self-assessment (Black & Wiliam 1998a, b). In the same way, cooperative learning can enhance self- and peer assessment (Crooks 1988).

Other factors for effective self-assessment are the following: students have to hold a concept of quality similar to that held by teachers, they also have to monitor the quality of their performances while performing them and they have to have alternative strategies. That means that students need to know the goal, which should be explained by the teacher (Sadler 1989; Harlen & James 1997; Black & Wiliam 1998a, b; Nitko 1989; Harlen & Winter 2004; Shepard 2005; Harrison & Howard 2009), compare the actual level of performance with that standard, and engage in appropriate action to close the gap (Sadler 1989: 121; see also Black & Wiliam 1998a, b; Harrison & Howard 2009). For students to hold a similar concept of quality, teachers have to share theirs with them. Sometimes this is not easy, as quality is often unarticulated and constitutes tacit knowledge in teachers' heads (Sadler 1989). If teachers do not share their guild knowledge, learners will always depend on them and their judgements to know about the quality of their performance (Sadler 1989). For students to compare their actual level with the reference level they have to make objective and detached judgements about themselves (Sadler 1989). Self-assessment, then, involves collaboration between teachers and students (Harlen & James 1997; Shepard 2005), although teachers seem reluctant to (Fontana & Fernandes 1994; see Black & Harrison 2001a for reports of how the transfer of responsibility for learning that AfL practices entail is not easy for all teachers). In spite of the importance of self-assessment, it seems students are not offered enough chances for making evaluative judgements (Sadler 1989: 140).

In spite of the difficulties, very positive results have been reported in the implementation of self-assessment. White and Frederickson (2000) demonstrated that judgements from students trained in self-assessment strongly correlated with teachers' judgements. Also, self-assessment has positive impacts on meta-cognitive skills, motivation, self-esteem, and learning (Black & Wiliam 1998b; Ollina & Sullivan 2004; Fontana & Fernández 1994; Crooks 1988; Shinn &

Hubbard 1992; Shepard 2005). The positive impact of self-assessment on students' meta-cognitive skills seems to be especially relevant, since these skills allow students to access cognition and learning, and to know their own strengths and weaknesses. In the end, as mentioned above, by fostering self-assessment, teachers would be promoting transfer of responsibility for the learning process.

There are different ways in which self-assessment can be implemented in AfL classes. One way of implementing self-assessment is giving out traffic lights to students, green meaning they understood or they did a good job, orange meaning they understood or did the task so-so, and red meaning they did not understand or they did not do well. The same technique can be implemented with thumbs up-thumbs down: thumb up would equal the green traffic light, thumb down the red one, and thumb not up nor down would mean the same as the orange traffic light (Black et al. 2004; Leahy et al. 2005; Wiliam et al. 2004; Black & Harrison 2001b; Harrison & Howard 2009). An advantage of this method is that teachers can notice the difficulties with just a glance (Black & Harrison 2001b). Giving students simplified versions of rubrics is another possibility of self-assessment (Black et al. 2004). WALT (what We Are Learning Today) and WILF (What am I Looking For) are two of the most widespread techniques to share and discuss criteria with learners, which is also essential for students' self-assessment (Harrison & Howard 2009). Leahy et al. (2005) report that the teacher they worked with said that students' self-assessments tended to be accurate and that students affirmed that self-assessment helped them improve their learning. Black et al. (2004) claim that it has been proved that students who used self-assessment for examinations and were trained by creating their own questions and answers, outperformed those students who were trained in the traditional way (Foos et al. 1994, as cited in Black & Harrison 2001b: 10).

Associated with self-assessment, peer-assessment is another important characteristic of AfL, as it also brings about benefits that improve students' learning processes (Black & Wiliam 1998 a, b; Wiliam et al. 2004; Ruiz-Primo & Furtak 2007; Rea-Dickins 2001). Peer-assessment involves students assessing other students' work, but not giving them grades (even less so grades that will be reported officially), focusing on learning and improving instead (Leahy et al. 2005), and not on competition (Black et al. 2003; Harlen & Winter 2004). Self- and peer-assessment should go hand in hand, as peer assessment helps develop self-assessment and as both involve students making use of meta-cognitive processes (Harrison & Howard 2009; Heritage 2010).

Peer-assessment has some advantages over self-assessment and over teacher assessment. Over self-assessment, because it might be easier for students to see their classmates' mistakes than their own mistakes (Leahy et al. 2005; Sadler 1989). If they assess a classmate's work, they are

less emotionally involved than when they assess their own work (Leahy et al. 2005; Sadler 1989; Wiliam 2006), although very young learners can be influenced by who they are friends with (McKay 2006). The advantage of peer-assessment over teacher assessment is that learners seem to share a code which leads to more effective communication among them, even when they take the role of the teacher (Sadler 1989; Leahy et al. 2005; Black et al. 2003; Rawlins 2010 to see how students prefer to ask their classmates for help instead of the teacher), and to accept criticisms in a better way (Black et al. 2003).

There are two main benefits for students when doing peer-assessment, the latter of which also applies for self-assessment. The first one is that this type of assessment can have benefits not only for the recipient of the assessment, but also for the student assessing, as they are forced to take the role of teachers and engage with understanding criteria and goals (Leahy et al. 2005; Sadler 1989; Heritage 2010; Wiliam 2006; Harlen & Winter 2004; Black et al. 2003). At the same time, what students say in the process of peer-assessing can be more evidence for the teacher regarding students' understanding of learning goals and success criteria, and can help them prepare adequate interventions (Heritage 2010; Black et al. 2004).

After thoroughly scrutinizing AfL and its main characteristics, and moving closer to the context of the present study, the next sections will address the role of AfL in specific content subjects, and in CLIL (where content subjects and a second language are learnt in integration).

3.2.7 AfL in different subject areas

Research in AfL has addressed generic strategies applicable to all subject areas, although they have been carried out in the context of specific subjects, mostly Mathematics and English (Fontana & Fernandes 1994; Craven et al. 1991; Hodgen & Marshall 2005). Black and Wiliam (1998a) and Hodgen and Marshall (2005) suggest that professional development needs to be done in different subjects. Although there are some general principles of AfL which apply to all areas, the way in which they manifest themselves in different subjects might be different, which is why establishing a community of subject discourse is necessary (Black et al. 2004; Wiliam 2006; Black & Wiliam 1998a; Black & Wiliam 2009).

Some subjects like Mathematics, Science, and English (as a first language) have received attention in AfL research (Harlen & Winter 2004). Some AfL studies have found differences among subjects. For example, Black et al. (2004) found differences when they compared teachers of Mathematics, Science, and Language Arts. For example, regarding assessment tasks, in the case of Science and Mathematics, there are features for which a correct model or explanation is offered, which often differs from the preconceptions students have. However,

more open tasks, such as opening up a discussion in which students' ideas are challenged and scientific models are provided might be more effective (Black et al. 2004). In other words, we have to differentiate between open tasks and closed assessment tasks (Black et al. 2004: 10). In closed tasks, there is just one right solution or answer, while in open tasks there is a range of possibilities. Black et al. (2004: 10) claim that tasks in Language Arts are normally more on the open side of the continuum, whereas in Mathematics they are more on the closed side.

Hodgen and Marshall (2005) also explored the differences in AfL between a Mathematics class and an English class. These two disciplines are thought to be opposites, but what made the two classes formative was the same: engage pupils in a task; extend pupils' thinking through justifications of their reasoning; and engage learners into peer assessment. The authors stressed that the teachers' profound subject knowledge enabled them to construct meaningful tasks and enact them.

All in all, AfL in different subjects areas is still an underresearched area. Only a few subjects have been studied and compared, these being mainly Mathematics and English. In the same way, only L1 content classes have been researched when subjects have been compared. The present study will contribute with other subject areas (Arts, Drama, and Citizenship) and new classroom contexts (CLIL).

3.2.8 AfL in CLIL

The little research that has been done on AfL in EAL/content-based/immersion classrooms (Davison & Leung 2009) reveals a lack of systematicity and variability because teachers' judgements are many times of an impressionistic nature (Breen et al. 1997; Davison & Williams 2002; Leung 1999; Leung & Teasdale 1997). This impressionistic nature is very much related to the fact that informal or interactive AfL is co-constructed and context-dependent, which means that teachers normally have seconds to interpret students' evidence and react to them (Black & Wiliam 1998b; Stiggins 2001; McNamara 2001; Brookhart 2003).

AfL is assumed to have more advantages than external exams, especially when it comes to assessing language, because to assess effective language development, both knowledge and skills and their applications in a wide range of different situations and modes of communication are to be assessed (Davison & Leung 2009: 401). Assessing language involves assessing content, especially in CLIL, so the question seems to be whether teachers need two separate assessment frameworks, or whether a content-language integrated view is possible (Davison & Leung 2009; Mohan, Leung & Slater 2010). Some researchers advocate for the latter, for "assessment of content requires a language-based theory of knowing and learning" (Byrnes

2008: 46-47). In this vein, Mohan et al. (2010) propose "Integrated Assessment of Language and Content" (IALC), which relies on SFL to assess language and content in integration: how wording constructs meaning. A functional approach to language can provide the tools to analyse and relate meaning and wording systematically in discourse, both written and spoken, as how meaning is achieved through wording is central for language assessment (Mohan et al. 2010: 225). Much of second language assessment seems to focus on a traditional view of grammar as rule, concentrating on errors, and failing to relate meaning-wording relationships (content-language) (Mohan et al. 2010: 230). However, content and language should not be assessed separately (Llinares et al. 2012; Mohan et al. 2010; Mohan & Slater 2004), and IALC is therefore not only possible but also necessary both in L1 and L2 environments. To implement it, teachers need support, recognition, and resources (Mohan et al. 2010). As an illustration, Slater and Mohan (2010) showed how texts can be functionally assessed in a systematic way for IALC through meaning-wording relationships. They key question of this approach is that teachers do not look at correctness anymore as traditional grammars do, but the focus is on what the learner can do (as opposed to what s/he cannot do). This approach drawing on SFL centres on discourse as a whole and how it works to convey the intended meaning that the content requires.

In relation to assessment, CLIL presents two main challenges: general educational assessment, and the assessment of foreign/second language proficiency. The primary focus of assessment in CLIL is content (Coyle et al. 2010). Therefore, it has more to do with general educational assessment (assessment of non-language subjects) than with second and foreign language assessment. Conventional language assessment does not meet the needs of CLIL, in which both CALP and BICS need to be assessed (Byrnes 2008; Llinares et al. 2012). In CLIL, assessment of content is influenced by students' performance and use of the foreign language; however, this linguistic element usually remains invisible (Hönig 2010). Therefore, it is important that the language required for the tasks is visible (see Nitko 1994 for a discussion on assessment aligned with curriculum learning goals). In Llinares et al.'s words (2012: 284), "[w]herever language is an invisible component of performance, as in many educational contexts, this can lead to gross unfairness in assessment practices".

Assessment is a key issue in CLIL, and particularly the question of the role that language plays in it (Llinares et al. 2012). Llinares et al. (2012) argue that AfL is essential if CLIL students are to meet both content and language goals. One of the advantages is that if issues of language are reserved for AfL, students' motivation is likely not undermined (Basse 2016), as they know that making linguistic mistakes will not be punished as long as they do not interfere with their expression of content knowledge or performance of skills. In turn, formal accuracy can be attended to when it is fundamental to understand and talk about the content (Llinares et al. 2012:

296). If AfL can be positive for the integration of content and language in CLIL classes, then teacher development in CLIL should include helping teachers understand how language can be used for AfL in different types of interactions (Llinares et al. 2012). In addition, teacher development in CLIL should also address the role that language plays in assessment, how to make it visible and clear for teachers, as it will affect students' content performance as well (Llinares et al. 2012: 284).

Assessment in CLIL should, then, reinforce both the learning of the subject and the learning of the language, and should not allow that the integration of language and content stands in the way of students' performance in tests of subject-specific skills and concepts (Dale & Tanner 2012: 38). That is, the integration of content and language should always favour learning and never detriment it.

There are three characteristics of AfL which, according to Llinares et al. (2012: 285-287), are especially important in CLIL: it is planned (they have to identify what they will assess and how they will collect the evidence), reactive (teachers need to modify their tactics based on the evidence they collect), and reciprocal (both teachers and students can benefit from it, adjusting their tactics according to the knowledge showed). These three characteristics are explained in more detail below.

1. AfL is planned

Teachers have to identify learning objectives and sequence them according to learning progressions (Llinares et al. 2012; Heritage 2008). How the evidence will be collected is another feature that needs planning: not only descriptions of concepts, knowledge and skills are needed, but also the language which accompanies them (Coyle et al. 2010). Genres and registers required need to be identified, along with the required grammar and vocabulary (Llinares et al. 2012). These language features are then reinforced during teaching and included in the summative assessment.

2. AfL is reactive

Instruction is constantly adjusting, using as a basis the evidence elicited about students' understandings (Llinares et al. 2012; Black & Wiliam 1998a, b). Sometimes, the instructional adjustments will need to focus on the language necessary to acquire content goals (Llinares et al. 2012). This will depend on the previous planning stage: the teacher may have made learning progression of linguistic elements and, in this way, s/he will know to what extent learners are able to make word-meaning relationships in order to understand content (Llinares et al. 2012).

3. AfL is reciprocal

AfL can help students improve those areas in which they are weaker. AfL encourages, in this way, learner autonomy, something which is linked to development of L2 competence (Llinares et al. 2012).

Marzano (2010) identified two main uses of AfL: scoring systems (keeping a record of scores to track students' progress over a whole year or over a teaching unit; summative marks can be based on these formative scores), and instructional feedback (through on-the-fly interaction in the classroom, meant for students to improve their performance; it is not recorded – this would equal to what other authors call informal, interactive, on-the-fly AfL, see above). Llinares et al. (2012: 10) claim that it is precisely through these two uses of AfL that the unique nature of CLIL is reflected, as register, genre, classroom interaction, scaffolding, and content knowledge are integrated. Based on Marzano's assessment scale (2010), Llinares et al. (2012: 287) propose a "content-language integrated assessment". Marzano's scale is based on different levels of complexity in achieving learning objectives. Her scale integrates content and language in an AfL framework. It includes a wide range of content classrooms and grade levels, but not CLIL. That is why Llinares, Morton and Whittaker (2012) add the language dimension to integrate both aspects (linking subject-specific genres and registers to cognitive complexity of content). Llinares et al.'s (2012) integrative tool is also based on Polias (2003), who details an SFL approach on language development for ESL learners and native speakers). The language needed for achieving learning goals needs to be explicit at various levels of complexity (Llinares et al. 2012: 298). Otherwise, teachers will not be able to provide useful feedback to students (Mohan et al. 2010: 230-231). This can be done precisely thanks to SFL, as genres are constructed through lexico-grammatical features (Llinares et al. 2012: 19). With this scale, which integrates content and language, teachers can give just one and appropriate mark for both content and language (Llinares et al. 2012).

Imperative things to improve CLIL programmes are that teachers be able to state language objectives (which go hand in hand with content ones), and to provide content-language integrated feedback and instruction to be included in professional development programmes (Llinares et al. 2012). As for integrated feedback, Llinares et al. (2012: 287) further stress that instructional feedback should focus on the linguistic features through which content is expressed and that CLIL teachers should have the linguistic skill of giving content-focused instructional feedback as part of AfL. The present study will deal with instructional feedback and how teachers focus on either content, language, or both in an integrated way in classroom discourse. For this purpose, the final section of this chapter addresses brings together sections 1 (classroom discourse) and section 2 (AfL).

3.3 CLASSROOM DISCOURSE AND AFL

This final section will focus on the important role of interaction in AfL.

3.3.1 Interaction and AfL

According to Black and Wiliam (1998a: 52), assessment can only be useful and effective if it happens in interaction. Meaning is constructed through interaction and all the interlocutors have a joint responsibility for the creation of meaning, identities, and events (Jacoby & Ochs 1995: 177; see also Rea-Dickins 2001; Heritage 2010; Black & Wiliam 1998b; Leung & Mohan 2004; Leung 2004; Davison & Leung 2009). In interaction, the teacher can respond to and reorient pupil's thinking, which is key in AfL. Ways to respond include opportunities for improvement in students' knowledge or inhibiting students' opportunities to learn if the teacher is constantly looking for the right answer, and allowing no flexibility to deal with the unexpected (Black & Wiliam 1998b: 143).

The teacher is a kind of mediator between a body of knowledge and skills to be learned and the learner (Black & Wiliam 1998a; Sadler 1998; Mortimer & Scott 2003; Barnes 1975; Davison & Leung 2009; Harrison & Howard 2009; Llinares et al. 2012). Thus, they need to know how to build bridges between the two entities. As Black and Wiliam put it (1998a: 16):

> All [classroom] work involves some degree of feedback between those taught and the teacher, and this is entailed in the quality of their interaction which is at the heart of pedagogy. The nature of these interactions between teachers and students, and of students with one another, will be key determinants for the outcomes of any changes.

For AfL to be successful, interaction "should be thoughtful, reflective, focused to evoke and explore understanding, and conducted so that all pupils have an opportunity to think and to express their ideas" (Black & Wiliam 1998b: 144). AfL allows students to be involved in interaction and to develop skills for meta-cognitive reflection about their learning (Rea-Dickins 2001: 452-453; see also Shepard 2005; Hodgen & Marshall 2005; Harrison & Howard 2009). In the same way, AfL allows students to connect concepts and skills from different subject areas creating new knowledge, which prevents compartmentalized knowledge and its detrimental effects from appearing (Shepard 2005). Dialogue, interaction, is evidence of learning, since students show their understandings and construct new ways of thinking (Harrison & Howard 2009: 8). In this way, paying attention to it can help teachers diagnose and plan future teaching steps (Harrison & Howard 2009: 7).

Interestingly, through interaction, a lot of the classroom time can be dedicated to assessment (AfL). In fact, in classrooms more time is spent on informal AfL than on tests (Rea-Dickins 2001: 434). Opportunities for informal assessment are embedded in good classroom practice (questioning, interaction, feedback, etc.), which sometimes becomes indistinguishable from AfL (Rea-Dickins 2001: 457). According to Leung (2004: 21-22), AfL is contingent, dynamic, dialogic, interactive, and co-constructed (see also Heritage 2010; Torrance & Pryor 2001; Ruiz-Primo & Furtak 2007; Bell & Cowie 2001; Black & Wiliam 2009). That is the reason why AfL cannot be accommodated into pre-specified criteria (Leung 2004: 23). Its contingent nature makes the same teacher change in different contexts (subjects) or with different students, since they do not invoke the same pedagogic assumptions or assessment criteria (Leung 2004: 23; Torrance & Pryor 1998). AfL is concerned with the creation and capitalization upon moments of contingency in instruction to regulate learning processes, be it real time adjustments during interaction (synchronous), or teacher's feedback through grading practices and evidence derived from homework (asynchronous) (Black & Wiliam 2009: 10). Thus, AfL takes into consideration "the interactive and contingent nature of student performance in the classroom which is dynamic and co-produced with the teachers and others" (Leung 2004: 22). Responsiveness, then, is one of the characteristics of AfL, it is key in any subject and it involves both planning and flexibility (Ruiz-Primo & Furtak 2007; Bell & Cowie 2001; Tharp & Gallimore 1988). A formative interaction is defined as one "in which an interactive situation influences cognition, i.e., it is an interaction between external stimulus and feedback, and internal production by the individual learner" (Black & Wiliam 2009: 11). The contingency of formative interaction makes the teacher's work unpredictable: s/he focuses on what s/he can learn about student thinking through their response (interpretive listening, Davis 1997). However, sometimes, what the teacher interprets is not what the learner intended to express or vice versa (Black & Wiliam 2009: 12). In formative interactions, teachers bring to bear prior knowledge taken from prior judgements of students' efforts on similar tasks (Black & Wiliam 2009: 13). Thus, a detailed analysis of teachers' moves would be needed for a better understanding of formative dialogue (Black & Wiliam 2009: 25).

The link between interaction, learning, and AfL is nicely illustrated in the following quotation from Bell and Cowie (2001):

> It is through the teacher-student interactions during learning activities (Newman, Griffin & Cole 1989) that formative assessment is done and that students receive feedback on what they know, understand, and can do. It is also in these student-teacher interactions during learning activities that teachers and students are able to generate opportunities for furthering the students' understanding. As formative assessment is viewed as occurring

> within the interaction between the teacher and student(s), it is at the intersection of teaching and learning (Gipps 1994). In this way, teaching, learning, and assessment are integrated in the curriculum (Bell & Cowie 2001: 539).

Therefore, "the socially co-constructed nature of formative teacher assessment therefore makes it necessary to attend to classroom interaction and classroom discourse as a key site for empirical investigation" (Leung 2004: 29; see also Torrance and Pryor 1995; Van Lier 1988). This entails a focus on how formative aspects are actually accomplished in the classroom (Leung 2004: 29). In teacher-student interaction, teachers have more power than students in planning, decision-making, etc., and they shape students' talk (Leung & Mohan in press: 339). Nonetheless, that does not mean that students are just passive agents, as they take their prior knowledge to the classroom and their interpretations of what is to be done in certain situations.

Investigations on interaction and AfL in actual classrooms are very much needed (Leung & Mohan in press; Leung & Lewkowicz 2006), for not all teachers are capable of distinguishing between AfL and summative assessment, and are ready to leave their preference for summative assessments behind (Broadfoot & Black 2004). Underlying these issues is teacher's professional knowledge and skills, and what educational values inform their everyday practice (Leung & Lewkowicz 2006: 228). Assessment through interaction has been addressed, among others, by Leung and Mohan (in press), Ruiz-Primo and Furtak (2006, 2007), Edelenbos and Kubanek-German (2004), Harlen and Winter (2004), Andrews (2004) in her review of Torrance and Pryor (1998), Torrance and Pryor (2001), and Rea-Dickins (2001). There is still much to do with respect to assessment as a social practice, and discourse analysis and ethnographic approaches are both important to this respect (Rea-Dickins 2004). Torrance and Pryor (1998) concluded that, to improve the quality of interaction and the effectiveness of AfL on learning, attention to the social construction and accomplishment of classroom assessment are needed. Leung and Lewkowicz (2006: 228) suggest that work in testing and assessment is greatly relevant to language pedagogy and curriculum development, and vice versa. This is particularly relevant in CLIL contexts, as quality AfL can lead not only to students' and teachers' better monitoring of the learning process, but due to the key role of interaction, students will also be able to use the second language not only to express content knowledge but also to reflect about their own learning.

Some of the most complete studies on interaction and AfL are Ruiz-Primo and Furtak (2006, 2007). Working with informal AfL practices in three middle school Science classrooms, they developed a model to analyse assessment conversations (Duschl & Gitomer 1997), or teacher-student interactions. Through questions, teachers try to use diagnostic strategies. The responses

to these questions are the evidence of students' current perceptions and they form the base on which teachers start working. The third step of the model is the teacher recognising the student response, and finally, the teacher acting upon the response. These are the three stages of assessment conversations (Duschl & Gitomer 1997). Ruiz-Primo and Furtak (2006, 2007) called these cycles ESRU cycles (Eliciting-Student response-Recognition-Using), and they defended they were distinct from IRF patterns in a number of ways. Mainly, the third move is not a mere evaluation, but an action for the students to take to improve their learning, such as "ask[ing] another question that challenges or redirects the students' thinking; […] promot[ing] the exploration and contrast of students' ideas; make connections between new ideas and familiar ones" (Ruiz-Primo & Furtak 2007: 20).

Ruiz-Primo and Furtak (2006) argue that more than one iteration of the cycle might be needed to reach a complete and appropriate understanding, depending on the nature of the conversation. In fact, in their study, it was proven that students with teachers who completed more cycles (higher quality ESRU cycles) during their classrooms performed better in tasks and exams (Ruiz-Primo & Furtak 2007). They also argue that the more complete ESRU cycles are, the more probable it is that the information gathered is helpful for learning purposes. Many times, several incomplete cycles were observed before the using move came. That was because the teacher either elicited responses from several students, or several responses from the same student.

As for questions, low percentages were found for eliciting questions asking students to relate evidence to explanations, evaluate quality of evidence, promote argumentation, or compare and contrast others' ideas (Ruiz-Primo & Furtak 2006, 2007). Of the questions asked in the conceptual domain, the most common were those that asked students for definitions (5%). A lot of yes/no questions and questions to check student understanding were found too (Ruiz-Primo & Furtak 2007). Likewise, the most common recognizing strategy was revoicing or rephrasing student responses. Sometimes, this was accompanied by elaborating on the response. Revoicing is considered a good strategy to engage students (O'Connor & Michaels 1993; Nystrand & Gamoran 1991), because it not only recognizes what the student is saying, but also constitutes an evaluation strategy, as the teacher elaborates on what the student has said. Equally, Nystrand and Gamoran (1991) found this strategy to have positive effects on achievement. The most frequent using strategy was why and how questions, because they redirect and challenge students' thinking. These types of questions provide the teacher with more information about students' understandings. Ruiz-Primo and Furtak (2006) also observed that it was very infrequent that teachers made criteria explicit or asked students to relate evidence to explanations. Meta-cognitive questions (questions such as "why do you think so?" or "what

does that mean?") are very good at eliciting students' understanding, but they are even more helpful when they are combined with other strategies, such as comparing and contrasting students' ideas and explanations (Ruiz-Primo & Furtak 2006). In general, the cycles were formed by long teacher turns and short student ones (Ruiz-Primo & Furtak 2007). Helpful feedback was not provided by teachers very frequently (Ruiz-Primo & Furtak 2006: 12). Asking questions and then simply repeating students' words was very common, as well as the use of implicit questions (called on a student instead of repeating the original question) (Ruiz-Primo & Furtak 2006). Ruiz-Primo and Furtak (2006) further state that "[t]he teacher whose students had the highest performance on our tests was the teacher who held the most discussions, asked the most concept-eliciting questions, and employed the greatest diversity of strategies that used information she had gained about student understanding" (Ruiz-Primo & Furtak 2006: 27). Collecting information about students' understanding can be done during any interaction and increases in students' performance can be linked to high-quality informal assessment practices (Ruiz-Primo & Furtak 2006: 28). In a similar line, Taasoobshirazi et al. (2006) claimed that it is possible to raise high-stake achievement scores while promoting inquiry-oriented learning.

The next two sections specifically focus on the role of questions and feedback on students' responses in AfL.

3.3.2 Questioning and AfL

Teacher questioning is considered one of the main and most important AfL techniques (Black et al. 2003; Wiliam et al. 2004; Black et al. 2004; Torrance & Pryor 1998; Black & Wiliam 1998a, b; Harlen & Winter 2004; Harrison & Howard 2009). Stiggins et al. (1989) show the concern about teachers' questions and their tendency towards asking more of the recall type than higher order types (see also Slavin 1991), even for the teachers trained to ask high-order questions. High-order questions are those which challenge children's perceptions of the world, reconstructing their understanding (Harrison & Howard 2009). Bloom's new taxonomy (Anderson et al. 2001) divides questions focusing on: remembering; understanding (describe, explain, paraphrase) (these two types triggering lower-order thinking skills); applying (demonstrate, illustrate, dramatize); analysing (compare, contrast, criticize, test); evaluating (argue, judge, evaluate); creating (construct, create, design) (these types triggering higher-order thinking skills). Using Bloom's new taxonomy (Anderson et al. 2001), Wragg and Brown (2001) found that 70–80% of classroom questioning focuses on low-order skills, such as remembering and understanding. In general, different studies have concluded that most teachers' questions have more to do with managerial aspects of the class and the recalling of

facts than with high-order questioning (Wragg & Brown 2001; Wragg 1993; Kerry 1989; Galton, Croll & Simon 1980).

Although richer questioning styles can have benefits, they are found to be difficult to implement in real classrooms (Black & Wiliam 1998a: 56; Dassa 1990). A question on its own is not formative (Black & Wiliam 2009; Harrison & Howard 2009; Black 2012). Students should be encouraged to respond, and these responses should be used, in turn, to explore students' ideas, create a dialogue, and move learners on (Black 2012; Harrison & Howard 2009; Black & Wiliam 2009). In other words, teachers should ask questions that realize scaffolding, questions that guide learning and elicit understanding (Torrance & Pryor 2001). Black, however, explains how teachers might use open questions but then assess the responses as correct or incorrect, which is contradictory (Black 2012). It is the feedback part which shapes students' understandings (Harrison & Howard 2009). However, teachers might be afraid that following students' contributions and ideas might lead the interaction into unpredictable paths, and thus, they will not be able to control the dialogue anymore (Black 2012).

Several AfL researchers insist on the importance for teachers to plan classroom questions in a way that it may help learning (Black et al. 2003; Black et al. 2004; Leahy et al. 2005). We know that most of the time teachers spend the time in class asking questions or in whole-class discussions. However, they have been criticized because they listen to students' responses evaluatively rather than interpretively (Davis 1997); that is, they listen to hear the correct answer and not to learn about students' thinking. According to Harrison and Howard (2009), some questions should aim at discovering where students are at; other questions should aim at creating new knowledge rather than rehearse existing one; others, which they called *hinge point*, aim at checking students' understandings to know where to direct the lesson. Leahy et al. (2005) propose some solutions in order to overcome two problems of traditional questioning in classrooms. The first problem is the hands-up policy. If students do not want to be engaged in the classroom, they simply do not put their hands up. However, if this policy is dropped and the teacher can ask any student, all of them must be alert and engaged in the instructions (Leahy et al. 2005; Black et al. 2004; Black et al. 2003; Black & Harrison 2001a; Harrison & Howard 2009). The second problem is the fact that teachers only get to know one student's thinking (Leahy et al. 2005; Furtak & Ruiz-Primo 2008). To hear all of them, all the students would have to respond to every question the teacher asks, and that is not feasible. That is why, in classes where AfL is implemented, students have whiteboards where they write their responses to teacher's questions and hold them up, and so the teacher can read them (Leahy et al. 2005; Harrison & Howard 2009). Other techniques are colour cards, lollipop sticks with the students' names on them, or collaborative peer talk before sharing the answer with the whole class

(Think-Pair-Share) (Harrison & Howard 2009). The aims of these strategies are to engage all students and prevent some of them from dominating the talk.

Another aspect that has been criticized by researchers is that, when the teacher asks a question, less than a second is waited on the part of the teacher to receive an answer (Black et al. 2004; Harlen & Winter 2004; Rowe 1974; Black & Harrison 2001a). After that time, if no answer is provided by the student, the teacher normally reformulates the question, answers the question him/herself, or asks another student. The consequence of this is that teachers normally ask questions for memorized facts, as only this type of question can prompt an answer almost without thinking, in less than a second. Then, the interaction between teacher and students is at a superficial level. Black et al. (2004), then, propose that teachers increase wait time. This is a way to challenge students, as they will see that the teacher is really waiting and that s/he really cares about what the learner has to say, forcing them to think and give an answer. This is a way of having more pupils engaged in the discussion and it may also have as a result longer contributions on the part of students.

For good questioning, then, teachers need to plan questions beforehand and according to learning purposes (Black & Wiliam 1998b; Leahy et al. 2005; Black et al. 2003), to anticipate the possible responses and the possible follow-ups for those responses (Black & Harrison 2001a; Black et al. 2003). Questions become, in this way, a fundamental tool for teachers, since they give them the opportunity of, first, exploring students' conceptions, and then, build on and develop them. Black et al. (2004) give the example of simple questions such as "Why do you think that?" as a very good way to enhance and extend students' thinking. For good questioning, Black et al. (2004) also advocate for a moving away from the classical factual questions. Abandoning factual questions means paying attention to quality and function of questions (Black & Harrison 2001a).

3.3.3 Students' responses and teachers' feedback

"Questions are only as good as the answers that they elicit" (Wragg & Brown 2001: 27). If we think that using language is important to language learning, and we think that linguistic processing is related to cognitive processing, then students' contributions are important (Van Lier 1996: 156). However, one needs to be careful with students' responses, as they may reveal what the student thinks the teacher wants them to say rather than their actual understanding or learning (Wragg & Brown 2001: 43).

Sadler (1989: 122) emphasizes that when AfL is implemented, it is more important to think of the quality of students' responses rather than valuing how many facts s/he remembers (facts

memorized; content or concepts mastered or acquired). A learner's response has to be understood in relation to their expectations and assumptions about classroom practice, as well as their interpretation of the task demand and the criteria for success (Black & Wiliam 1998a: 55; Aikenhead 1997).

Black et al. (2004: 9) say that, more important than a high qualification on a subject, is the understanding of its fundamental principles and the ability to anticipate the difficulties students might have, and to think about questions which stimulate thinking. These abilities will be essential to interpret responses, because in them there might be clues as to what misconceptions students have. Teachers have to interpret student responses (in terms of what they reveal about their understanding), and then decide on the best feedback, thus fulfilling some of the main characteristics of AfL, such as gathering and interpreting evidence, and deciding on the best course of action based on the evidence collected (Black & Wiliam 2009: 11). Teachers usually start a lesson with an opening move, which is typically a broad exploratory question to elicit students' conceptions about the topic. Students' contributions may raise new possibilities, and the teacher has to decide whether to follow them or not. Two opposing forces appear here: advancing the learning of the whole class, and avoiding making students feel rejected (Black & Wiliam 2009: 21). These forces may seem opposing because curricula are usually extensive and teachers are always pressured to finish them. Therefore, the time that can be devoted to students' spontaneous contributions is limited by those curricula.

Ruiz-Primo and Furtak (2007) suggest that student responses need to be longer than just one word answers and need to contribute to classroom conversation. In this sense, research says that the more students explain and justify their viewpoints, the longer their utterances will be (Mercer et al. 2004). However, several studies have demonstrated that students' contributions to classroom discourse tend to be rather short (Dalton-Puffer 2007; Cazden 2001; Barnes 1975). Other studies have found contradicting evidence (Black et al. 2003): classes in which students contributed with short phrases, and other classes in which students construct complete sentences (formative dialogue) and use *I think* and *because*, signals of reasoned thinking. Dillon's (1998) study is one of the few that have shown students talking as much as the teacher.

Improved practice would mean guidance to teachers when they interpret students' responses and to match their contingent responses to their learning intention. This might start with analyses of classroom interactions. Knowing the learning theories and specific issues involved in implementing a formative approach would help as well. It would be interesting to create a list of rules coming from extensive data, but it would have dangers: firstly, a response cannot be interpreted outside the context in which it was produced; secondly, a rule might apply in one

subject, or in one lesson aiming at specific goals within a subject, but might be inadequate for another subject or another lesson of the same subject with different learning aims (Black & Wiliam 2009: 27).

Within AfL literature, Wragg and Brown (2001) highlight different ways of giving feedback: answer/comment ignored (teacher asks someone else, changes question or topic); answer/comment acknowledged (teacher nods, smiles or says *yes*, *right*...); repeated verbatim (teacher re-states students' response and converts it into a question); part of an answer echoed (teacher re-states the acceptable part of student's answers and converts it into a question); praise contribution (teacher praises contribution, maybe with elaboration); corrected (teacher corrects incorrect part of an answer or asks others to correct); prompted (teacher asks prompting questions when first answers are inadequate giving hints to pupils); probe (follow-up questions when the first response was inappropriate but requiring more precise and thoughtful answers than prompts). This classification was important for the final feedback taxonomy of this study, as will be seen in the next chapter.

3.4 SUMMARY

This chapter has provided the main theoretical insights necessary to understand and frame the present study. The first section was devoted to classroom discourse, as it is one of the pillars for this research. In it, different topics have been addressed, such as the concept of interaction in general, scaffolding in interaction, IRF patterns, specific features of interaction in CLIL contexts, and the relationship between interaction and learning. This first section has specifically addressed the role of classroom interaction in language learning, making special emphasis on the Interaction Hypothesis and the importance of negotiation of meaning in SLA. A complete subsection has presented the different types of corrective feedback and their different implications for L2 learning and development. The second section of this chapter has been on Assessment for Learning, another pillar for this investigation. The most relevant characteristics of AfL have been fully presented and described, with special emphasis on some of them, such as feedback, self-assessment and peer assessment. Other aspects, such as practical and political considerations, types of AfL, and types of evidence have also been addressed. Especially important for this study is the relation between AfL and CLIL contexts and how the former can be helpful for the latter. The final section of this chapter has linked the two main theoretical pillars of this investigation: interaction and AfL. This section has stressed the major role that interaction plays when implementing AfL. In the same way, it has made evident why research studies that go into actual classrooms and analyse the interaction that goes on inside the black box are very much needed for the development of AfL theory and practice.

4

Data and methodology

This section will be devoted to the presentation of the data and methodology used for the analysis, also focusing on the difficulties found when analysing the data and on the decisions taken in order to solve those difficulties.

4.1 DATA

The data used for this research is part of a bigger corpus collected in the academic year 2010-2011. The corpus consists of 500,000 words over a total of 80 class sessions in 5 bilingual primary schools. In addition, there are also teacher and student interviews, and student surveys. For the purpose of this dissertation, 44 sessions out of the 80 total were selected from 4 different schools, which makes an approximate total of 300,000 words. The remaining 36 sessions were discarded, either because they were not CLIL sessions or because they corresponded to a different school level from the one under study.

Data from the 4 schools under analysis belong to two different bilingual projects existing in the Community of Madrid. These two projects are the MEC/British Council Project, which started in 1996, and the CAM Bilingual Project, which started in 2004 (see Chapter 2, section 2.2).

The four schools in which data were collected were grouped in two categories: *Non-AfL schools* and *AfL schools*. AfL schools are those in which Assessment for Learning is implemented and teachers have been trained for it, whereas Non-AfL schools are those in which Assessment for Learning is not specifically implemented. Training in AfL, although frequent in other countries like the UK, Hong Kong, or Australia, is very recent in Spain and is limited to a number of pilot CLIL schools within the MEC/British Council Project. In our data, the two schools belonging to the MEC/British Council Project are the ones with a specific implementation of AfL, whereas

the schools belonging to the CAM Project do not have any specific guidelines or training in AfL.

The data used for this study were collected in the third cycle of Primary Education: 5th and 6th year, in which students are, correspondingly, 10-11 and 11-12 years old.

Table 4.1 below illustrates the whole data set used in this thesis. The data were collected in four schools, two AfL schools (School 1 and School 2) and two Non-AfL schools (School 3 and School 4). From each school, one teacher collaborated in this project. Three of the schools were state schools (the two AfL Schools 1 & 2, and Non-AfL School 4), whereas School 3 (Non-AfL) was a semi-state school, meaning that it is partially funded by the State. As for the socioeconomic areas of the schools, middle-class districts predominate, although the zone of School 1 (AfL) could be better defined as an upper-middle class area. Two of the schools (AfL School 2 and Non-AfL School 3) were situated in the centre of the city, whereas the other two (AfL School 1 and Non-AfL School 4) were placed in the outskirts.

In the *Subject* column, there are all the school subjects represented in the corpus: Science, Arts, Citizenship, and Drama. The variety of subjects arises from the fact that different subjects are offered in English in the CLIL programme in different schools. In AfL School 1, the subjects were Citizenship and Science; and finally, AfL school 2, the subjects were Citizenship and Drama; in Non-AfL school 3, the subjects taught in English were Science and Arts; in Non-AfL school 4, the subjects were Citizenship and Arts.

			Subject	Year	Sessions	Time	Didactic units
AfL	School 1	Teacher 1	Citizenship	5	8	6h40	A circle of smiles The Giving Tree
			Science	6	8	6h35	Sound Bones and Muscles
	School 2	Teacher 2	Citizenship	5	4	3h20	Emotions Being healthy
			Drama	6	2	1h30	Word association Improvization
Non-AfL	School 3	Teacher 3	Science	5	7	6h	Vertebrates Pre-History
			Arts	6	4	3h20	Cubism Pop art
	School 4	Teacher 4	Citizenship	5	10	8h20	Democracy Gender
			Arts	6	4	3h20	The Alhambra Parallel lines

Table 4.1 Data set.

Following *school*, *teacher*, and *subject*, **Table 4.1** presents the school years which students were in at the time of data collection: 5th and 6th year of Primary Education. As shown in the table, in each school there is one group in 5th year Primary and another one in the 6th and last year of Primary Education. Next, the number of sessions that were recorded in each group and subject are given. As two complete didactic units were recorded in each group (each corresponding to a different subject), the number of sessions recorded in each subject varied depending on the length of the didactic unit. The last column of the table shows the total number of hours recorded in each group and subject (session length varied from 45 to 60 minutes).

Regarding the teachers, the AfL teacher in School 1 had 21 years of teaching experience, both at private and state schools, and in different countries (United States and Spain). The teacher from AfL School 2 had been teaching in state schools for 9-10 years, half of them in England, the other half in bilingual schools in Spain. As for Non-AfL teachers, the teacher from Non-AfL School 3 had eight years of teaching experience, three of them at a state school in the United States, three at a private school in Spain and two at a semi-state school in Madrid. Finally, the teacher from Non-AfL School 4 had a teaching experience of 10 years, 8 of which were in state schools and 7 specifically in a bilingual program. As for their command of the language, the two teachers from the AfL schools were native speakers of English with a good level of Spanish. The two teachers from the Non-AfL schools were native speakers of Spanish with an advanced (C1) level of English. All of them were both content and language specialists.

As illustrated in **Table 4.1**, the corpus includes a wide range of subjects recorded: Science, Arts, Citizenship, and Drama. Due to the fact that different schools offer different subjects in their CLIL programs, it was not possible to collect data on the same subject in all 4 schools. Teachers' availability to collaborate, as well as schools' willingness to go through the bureaucratic process of ethics in data collection, are key factors affecting the type of data used for educational research, and this is also the case in the present study. However, it was possible to record each teacher teaching two different subjects. This certainly enriches the corpus and the study, as it allows us to determine the role played by the subject in the classroom discourse generated by teachers and students, and thus, contribute to extend the scarce research on AfL as enacted in CLIL subject-specific discourses in primary education.

The collection of the corpus took a whole academic year because recordings were made at two different times (one at the beginning and one at the end of the school year). Complete didactic units were recorded at these two different times in each of the subjects and in each of the schools. During the recordings of the units, normal classroom activity in each unit was recorded for the purpose of the analysis.

All the classroom sessions were video-recorded, with the previous consent of the school and the students' parents. The posterior transcription of the data was done following the Santa Barbara papers (Du Bois et al. 1992), as the resulting transcripts are reader-friendly, and at the same time, include all the relevant information.

4.2 A MODEL FOR THE ANALYSIS OF AFL

This section of the dissertation will thoroughly explain the methodology adopted in the study. Specifically, the methodology of the analysis will be divided into five subsections: episodes, teachers' initiations, students' responses, students' initiations, and teachers' feedback. The first layer of analysis was the classification of the sessions into episodes. Then, those episodes selected were analysed in another layer: IRF patterns (in which the focus was on teachers' initiations, students' responses, students' initiations, and teachers' feedback).

All the data analysis (codification) and the statistical calculations carried out in this thesis have been done using the software *UAM Corpus Tool* (O'Donnell 2008). To ensure reliability of the data analysis, representative samples of the corpus were also independently categorized by a second researcher. After discussion on disagreements and on the most problematic cases, the Cohen-Kappa coefficient reached 0.80.

4.2.1 Episodes

Firstly, every classroom session was divided into episodes. Due to the large amount of data used for the present study, the division of each session into episodes required selecting the relevant episodes for further analysis. This was also done in previous studies (see Frölich, Spada & Allen 1985; Bloome et al. 2009; Snell & Lefstein 2011; Berg 2009). However, the majority of the sessions were analysed almost entirely, as most episodes were considered relevant. Next, **Table 4.2** enumerates the different types of episodes identified, and which were retained/discarded for further analysis. This categorization of episodes was open-ended, that is, it was based on the characteristics of the data and did not use any predetermined prescriptors (Frölich, Spada & Allen 1985). Thus, each category was named on the basis of what teacher and students were doing. However, some types of episodes were based on Snell and Lefstein's classification (2011), namely, *explanation of activity or homework*, *introduction of topic*, *revision of lesson*, *group/pair discussion/work*, *whole-class discussion*, and *group/individual student presentation of work.*

Type of episode	Analysis		
	Yes	*Partial*	*No*
Revision of lesson (last and current)	+		
Introduction of topic	+		
Whole-class discussion	+		
Group/individual student presentations of their work	+		
Evaluation/Correction of activity/homework	+		
Self- and peer-assessment	+		
Stating objectives for the lesson		+	
Students doing activity individually		+	
Group/pair discussion/work		+	
Teacher explanation of activity or homework		+	
Explaining marks		+	
Reading from the book			+
Listening from the book			+
Dictation			+
Singing song(s)			+
Classroom management/class routines			+

Table 4.2 Classification of episodes and their analysis.

The main criteria used in order to include or discard episodes for their posterior analysis was that of meaningful interaction: whenever there was interaction between teacher and multiple students, teacher and one student, or between students about content, language, or both, that episode or part of the episode was analysed. There were episodes that were entirely analysed, as the above mentioned criteria were fulfilled during the whole episode. These were *revision of lesson*, *introduction of topic*, *whole-class discussion*, *groups/individual students' presentation of their work*, *evaluation/correction of an activity or of homework*, and *self- and peer-assessment*. Other episodes were partially analysed. Only those parts of the episode in which meaningful interaction on content, language, or both occurred were further examined. The partially analysed episodes were, then, *stating objectives for the lesson* (for example, the part when students were reading aloud the objectives was not included in the analysis), *students doing an activity individually* (if students interacted with the teacher asking doubts about content or language, it was analysed), *the activity of group/pair discussion/work* (meaningful interaction in this episode occurred when the teacher approached the group or pair asking them questions to help them and when students ask questions themselves to the teacher; interaction between pairs or among groups with no teacher participation was not taken into account), *explanation of activity/homework*, and *explaining marks* (for these two last episodes, if IRFs did not arise and there were only teacher monologues, they were not analysed; if, on the contrary, student participation gave rise to IRFs, those parts were examined). The only episodes that were completely discarded for the analysis were: *reading from the book* (when students were reading one by one, at the teacher's call, a text from the book); *listening* (students listening to a CD); *dictation* (students copying what the teacher was dictating); *singing songs*; *teacher assigning*

homework for next class (because students did not intervene in discourse); and *classroom management and routines* (getting students organized, getting students into groups, greeting and stating the date, scolding students when they are not paying attention…).

4.2.2 Teachers′ initiations

Within teachers' initiations, the focus of this part of the analysis will be on teachers' questions. The definition of question that I will be following is any expression "used to elicit information or a response" (Crystal 1991:287). This definition does not only include interrogatives but also "linguistic commands", defined by Riesco Bernier (2007) as those commands that require a verbal response.

For the classification of teachers' questions, I have adapted Dalton-Puffer's (2007) typology of questions based on academic functions:

Question for facts	asking for objective happenings, something which is known to have happened or to exist
Question for explanations	asking for how something happened, and for elaboration of facts
Question for reasons	looking for reasons, arguments or causes why something happened
Question for opinions	asking for the students' personal opinion about a fact or issue.
Meta-cognitive question	asking students to argue a viewpoint or to articulate their thinking, making them aware of their own mental processes

Table 4.3 Dalton-Puffer's (2007) typology of questions.

The decision to adopt Dalton-Puffer's typology was based on a number of reasons. First of all, this classification focuses on academic functions and goes beyond the well-known dichotomies of open/closed or display/referential types of questions. Since the context of the study are CLIL classrooms, where both content and language are learnt, it seemed relevant to use a typology addressing academic functions. In addition, as Dalton-Puffer (2007) used this typology in Austrian CLIL classrooms, it would be interesting to apply it in a different context and observe possible similarities and differences between Spanish and Austrian CLIL classrooms.

In the adoption of Dalton-Puffer's typology, some changes needed to be introduced in the original taxonomy in order to satisfy the needs of the present data. The changes introduced were, then, data-driven. These modifications were the following: the type *questions for facts* also included questions for definitions (quite frequent in the present corpus); in addition, within the type *questions for explanations*, questions for *elaboration* were also included in this study, comprising those questions which encourage learners to elaborate on a topic or idea.

The following extracts from the corpus illustrate the different types of questions analysed (instances of question types are underlined):

Extract 4.1 Question for facts/definitions

TCH: [So] we are going to review which are the four... or five.. different eh vertebrate groups, ok? So, who can tell me which are the four or five vertebrates groups? Miguel
STU: eh amphibians, mammals
TCH: Ok, wait a second. Amphibians

Extract 4.2 Question for explanations/elaborations

TCH: Neolithic. Ok, so characteristics of people who lived in the Palaeolithic period. […] How did they eat? Inés?
STU: they hunt animals
TCH: They hunted animals, no? For food

Extract 4.3 Question for reasons

TCH: Why is it not so healthy to eat ice cream? Beatriz?
STU: Because it's like a bomb to your stomach. Because it's so, em, cold

Extract 4.4 Question for opinions

TCH: What is the most difficult in your opinion?
STU: A

Extract 4.5 Meta-cognitive question

TCH: Ok ((puts in James Brown's song "I feel good")) Your eyes are jumping out of your face! Why? [2]
STU: Because, I don't think that this song is going to be help us.

In order to account for all the examples in my data, in addition to Dalton-Puffer's typology, new types of questions were introduced. These were *language questions* and *meta-questions*.

The category of *language questions* defines all those questions that ask about specific linguistic aspects. It was added to the taxonomy as I was interested in how language was integrated in content classrooms. Following Snow et al. (1992) and their classification of language objectives, language questions were classified into *content-obligatory* or *content-related*. Content-obligatory questions are those which deal with the language required for students to

[2] It was sometimes difficult to decide when "why" introduced questions for reasons or meta-cognitive questions. The difference is that questions for reasons refer to content in a more objective way whereas meta-cognitive questions are more related to students' personal way of thinking and of arguing about a certain idea.

"develop, master and communicate about a given content material" (Snow et al. 1992: 30). For every topic, there are certain linguistic features which are essential for understanding and talking about the material. Content-related language questions refer to other language skills that can be compatible with the concept being taught. That is, they "*can* be taught within the context of a given content but *are not required* for successful content mastery" (Snow et al. 1992: 31). This distinction allows us to align CLIL and formative assessment pedagogy: how important is the language aspect for students to learn the content? In line with Llinares et al. (2012), if language objectives are identified and made visible, then the language necessary to understand the content (content-obligatory) would also be accessible and stressed. **Extracts 4.6** and **4.7** illustrate the two types of language questions: content-obligatory and content-related, respectively.

Extract 4.6 Content-obligatory language question

TCH: Copper is? Copper. Which metal is copper?
STU: <L1 ¿Cobre? L1>
TCH: <L1 Cobre L1>, and? Bronze. Which metal is bronze?
STU: ((Many)) <L1 Bronce L1>

Extract 4.7 Content-related language question

STU: Does. But here we have "does a joint help" "does the skeleton have" so maybe you can't see this next question, this is "does the skeleton have" so let's imagine you just see this question, "Does a joint…." Is this word "help" or "helps"?
STU: Help

In **Extract 4.6**, the teacher is asking for the translation of copper and bronze, two metals which are indispensable for students to learn the concepts of the unit on Prehistory. In **Extract 4.7**, on the other hand, when explaining bones and muscles, the teacher takes advantage of the context and asks about the third person singular *–s* of the present simple. Therefore, it is not a crucial linguistic aspect to learn the content of the unit, but the interaction allows the teacher to emphasize that specific grammatical aspect.

Meta-questions are questions that make students reflect on their learning (their improvements, their weak areas) and/or about assessment (promoting self-assessment), or questions that make students assess other classmates' work (promoting peer-assessment). Prima facie, this type of questions is hypothesized to be only found in AfL schools, since they represent a very important element of Formative Assessment: students are engaged in the assessment and learning processes, being active and increasingly autonomous agents (Black & Wiliam 1998a, b; Rea-Dickins 2001; Bell & Cowie 2001; Davison & Leung

2009). As a basis for this type of questions, it is important to refer to Chamot and O'Malley's (1987) meta-cognitive strategies such as self-evaluation and self-monitoring as part of the Cognitive Academic Language Learning Approach (CALLA), a content-based approach which helped limited English proficient students be prepared to go into the mainstream classroom in the United States.

Extract 4.8 below provides an example of a meta-question, as the teacher is asking students to value their capacities, what they have learnt, and what they still have to work on for a more complete mastery. In **Extract 4.9**, the teacher is asking about the minimum students need to reach to get a <L1 sufi L1>. The teacher is, therefore, providing students with explicit criteria that can help them when doing both self- and peer-assessment. With those criteria, students are able to assess their classmates (see **Extract 4.10**).

Extract 4.8 Meta-question: make students reflect on their learning

TCH: Can you please look our WILF? ((The WILF is on the board)) Can you tell me if you think you'd be able to do that or do you think there's something that's a little bit unsure about? ((They raise their thumb))

Extract 4.9 Meta-question: promoting self-assessment

TCH: So what do you have to do to get a <L1 sufi L1>?

Extract 4.10 Meta-question: promoting peer-assessment

TCH: What is their mark? […] Is it a <L1 bien L1> or a <L1 notable L1>? Or one mark for one person and one mark for the other person?
STU: He's not a <L1 sobre L1> because he didn't make the gestures.

The final typology of teachers' types of questions (Dalton-Puffer's typology along with the new categories included in this research) is shown in **Table 4.4.**

Question for facts	asking for objective happenings, something which is known to have happened or to exist
Question for explanations	asking for how something happened, and for elaboration of facts
Question for reasons	looking for reasons, arguments or causes why something happened
Question for opinions	asking for the students' personal opinion about a fact or issue.
Meta-cognitive question	asking students to argue a viewpoint or to articulate their thinking, making them aware of their own mental processes
Language question	asking about specific linguistic aspects
Content-obligatory	asking about linguistic features which are essential for understanding and talking about the material
Content-related	asking about other language skills that can be compatible with the concept being taught but are not obligatory
Meta-question	Asking students to reflect on their learning

Table 4.4 Final taxonomy of teachers question types.

4.2.3 Students′ responses

Moving on to students' responses, the analysis will help us identify learning moments in interaction. However, this study will not focus on psychological processes of learning or notions such as long-term learning. Students' responses will be measured using objective measures (length and complexity). Adapting Frölich, Spada and Allen (1985) and their COLT scheme, and to keep the analysis as simple as possible, I will classify students' responses into *minimal responses*, *T-units* (either one-phrase, one-clause, or more-than-one-clause T-unit), and *truncated responses* (see **Table 4.5**).

Minimal response	yes/no
T-unit	main clause (+ subordinate clauses)
One-phrase T-unit	the only explicit element in the T-unit is a phrase
One-clause T-unit	response consists of a clause (independent or subordinate clause)
More than one clause T-unit	response consisting of one (or more) T-units having more than one clause
Truncated response	response consisting of an incomplete T-unit

Table 4.5 Taxonomy of student response types

Minimal responses are those responses which consist of just *yes* or *no*, as illustrated in **Extract 4.11**. These minimal responses could also be considered an elided T-unit ("yes, I wrote it down properly") and hence could be classified as a subcategory within the category of T-unit.

However, rather than giving minimal responses the status of subcategory, I chose the option of having a category of its own (as the COLT scheme has) in order to overtly contrast those occasions in which students engaged only minimally in interaction with those other moments in which learners were more interactionally engaged.

Extract 4.11 Minimal response

TCH: Who was allowed, difficult word, pay attention please. Allowed. Who was allowed. Did someone get it right? Did you write it down properly? Who
STU: Yes

T-units consist of a main clause plus any subordinate clauses attached to it, if any (Hunt 1965). The use of T-units as a measure of students' language complexity has been mostly applied to written discourse (Hunt 1965, Martín Úriz & Whittaker 2005), but some studies have also used T-units for the analysis of spoken data (Long 1983, Young & Milanovic 1992, Young 1995, Pica et al. 1987, Nippold *et al*. 2009, Klecan-Aker & López 1985). T-units were classified as follows:

One- phrase T-unit: when the only explicit (not elided) element in the T-unit was a phrase, either a one-word phrase (instantiated in **Extract 4.12**) or a multiple-word phrase (instantiated in **Extract 4.13**):

Extract 4.12 One-word phrase T-unit

TCH: And which is the new unit? Which one is the new unit that we are speaking about something new? Em, Mustafa?
STU 1: Em, I forget
TCH: You forgot, ok, Pedro?
STU 2: Democracy

Extract 4.13 Multiple-word phrase T-unit

TCH: What are we going to look at today in Citizenship? Ana?
STU: About the %x…x%
TCH: About…excuse me?
STU: About the health

One-clause T-unit: when the response consists of a clause (independent or subordinate clause). **Extract 4.14** illustrates this category:

Extract 4.14 One-clause T-unit

TCH: Ok, so people now are following directions. I have blue eyes is wrong. Why is it wrong? Rocío, why is it wrong?
STU: Because we are talking about out perso- personality

More than one-clause T-unit: responses consisting of one (or more) T-units having more than one clause. This case is illustrated in **Extract 4.15**:

Extract 4.15 More than one-clause T-unit

TCH: Ricardo, what are you writing? I didn't say anything
STU: eh I am writing because in number five put answer the question on the whiteboard

The last category, truncated responses, has been considered because a response consisting of one incomplete T-unit could not be properly analysed within one of the T-unit types. They are illustrated in **Extract 4.16**:

Extract 4.16 Truncated response

TCH: Did they have gestures and movements?
STU: Yes
TCH: Yes, what did they do?
STU: Eh.. because he.. eh.. eh..

This approach to the analysis of students' responses will let us know what types of questions receive more complex answers and hence contribute more to students' language use and learning. In the same way, whenever there is uptake (students' response) after a teacher's feedback move, this analysis will allow me to know what types of feedback receive uptake or response and what type of response it is. It has been shown that student uptake and repair after teacher feedback provides evidence of the effectiveness of that type of feedback (Lyster & Ranta 1997). Likewise, the more complex the uptake is, the greater the evidence of learning will be.

4.2.4 Students' initiations

Students' initiations can take place with the previous teacher's consent (by giving them the floor) or without it. In their initiations, the learners may be asking a question (related to content

or language), stating a fact, offering an explanation, giving an opinion, arguing, or relating content to their personal experiences (see **Table 4.6**). This classification was based on the classification of teachers questions (see **Table 4.4**), although appropriate adjustments needed to be made in order to fit the data.

Asking a question
Content question
Language question
Stating a fact
Offering an explanation
Offering an opinion
Arguing
Relating content to personal experience

Table 4.6 Taxonomy of student initiations

Extract 4.17 below is an illustration of a student initiation, in which the student is giving his opinion and arguing about it, using arguments that have been already mentioned and explained in the class. Therefore, this student initiation would be an example of arguing, since it includes reasons and justifications.

Extract 4.17 Student initiation: offering opinion, arguing

TCH: Yes, it is. Watermelon is natural, it's a fruit, a lot, gives you lots of liquids, ok, and it cools you down. Can you…when it's hot you need to take things that make you cooler, to cool you down. Uh, Pedro??
STU: That I think that boy is very clever because he does not eat ice cream because it's hot and he eats a watermelon that is healthier.

For the purpose of this study, it was predicted that a more dialogic type of teaching (more aligned with AfL) will allow for more students' initiations. On the contrary, a more transmissional type of teaching will not be concerned about students' interpretations and experiences, thus allowing less space for students' initiations.

4.2.5 Teachers' feedback

Finally, teachers' feedback in both AfL and non AfL CLIL classrooms will be analysed. Many different categorizations of teacher feedback can be found in the literature, depending on the different research paradigms (see Lyster & Ranta 1997; Nassaji & Wells 2000; Frölich, Spada & Allen 1985; Sinclair & Coulthard 1975). The taxonomy applied in this study is based on the

previous literature on the topic and is also partly driven by the analysis of some of the data. Also, in an attempt to align with formative assessment theory, the categories are kept as simple as possible. The reason why formative assessment theory prefers simple categories is because it is an open-ended meaning-making system, and too many or too complex categories may present limitations, and thus, make the analysis somewhat artificial. In this line, some researchers (especially from the sociocultural paradigm) are against assigning a pre-established set of categories to third turns, arguing that in this way interpretations are limited and imposed by the researcher and different layers of meaning cannot be accessed (Lee 2007; see also Jarvis & Robinson 1997 and Gibbons 2003). Thus, in this study special attention has been paid to the fact that categories not only come from the previous literature but also emerge from the data. **Table 4.7** below shows the final taxonomy used in this study for the analysis of feedback.

Evaluation	
Positive evaluation	Evaluation recognizing the student's response as correct.
with recast (conversational)*	Reformulation of student's utterance, minus the error; followed by topic continuation (non-corrective repetition).
without recast	
Negative evaluation	Evaluation recognizing the student response as incorrect.
with explicit correction*	Provides the correct form of a previously incorrect utterance.
without explicit correction	
Expansion	Either asking further questions to push students to continue talking; or an extension of the content through the addition of related information.
Revision	Revise and/or summarize concepts and content
Re-route	Attempt to guide students to the right answer after a wrong response.
*Prompt**	Elicitations, repetitions, clarification requests, and metalinguistic feedback.
*Recast (didactic)**	Reformulation of student's utterance minus the error; provides explicit negative evidence intended to prompt the student to reformulate or self-repair, or enter a negotiation of meaning or form.
Meta-feedback	Either establishes success criteria and how pupils' work could be improved to meet those; or makes students reflect on their learning processes and possible future steps to take.

Note: * distinction between *focus on content* or *focus on language*

Table 4.7 Final taxonomy of teacher feedback types

The different paradigms I have based this feedback taxonomy on belong to the fields of second language acquisition (SLA), sociocultural theory, discourse analysis, and AfL. Within the SLA tradition, Lyster and Ranta (1997) analysed teacher feedback in French immersion classrooms. They classified feedback into explicit correction, recasts, elicitations, repetitions, clarification requests, and metalinguistic feedback (see Chapter 3, section 1.3.2). In brief, explicit correction provides the correct form of a previously incorrect utterance; recasts are reformulations of students' utterances made by teachers, minus the error; elicitations ask for reformulations of students' utterances by different means; in repetition, the teacher repeats the student's utterance with the appropriate intonation to highlight the error; clarification requests explicitly ask for clarification; and metalinguistic feedback is information or questions about the well-formedness of a student utterance. From these categories, I have used explicit correction as a subcategory of negative evaluation (teacher's evaluation recognizing the student response as incorrect); recasts as subcategories of positive evaluation (teacher's evaluation recognizing the student response as correct) and re-route; and what they called prompts, including elicitations, repetitions, clarification requests and metalinguistic feedback, among other things, as a subcategory of re-route.

Lyster and Ranta's categories created for French immersion classrooms were very adequate for the present data, probably because French immersion and CLIL are both content-based contexts and these categories can be applied to both content and language. At the same time, connections and relations between these categories and AfL can be easily established, for when prompts are used, as opposed to explicit correction, teachers react to students' misconceptions in a way that benefits their learning, guiding them in the quest for a more appropriate answer. Regarding *recasts*, and drawing on Lyster's (1998b) work on the ambiguity of recasts and students' failure to recognize certain types of recasts as negative evidence, in this study the two types of recasts are included under two different categories: evaluation and re-route, the former equivalent to conversational recasts, the latter to didactic recasts (Lyster 2007; Lyster, Saito & Sato 2013). Thus, the type of recast that is normally followed by topic continuation (with the function of non-corrective repetition) would be the one falling under the category positive evaluation in the present study, that is, teachers' evaluations of a student's response as correct (evaluating recasts). In turn, those recasts falling under re-route are the opportunities teachers attempt to create for students to reformulate or self-repair what they have said. Therefore, these recasts are expected to be noticed as negative evidence and to lead to a negotiation of meaning or form (re-routing recasts). Re-routing recasts are explicit and, therefore, more likely to be followed by uptake and repair (Llinares & Lyster 2014; Lyster 1998a). It is the teacher's goal for students to

notice that recast as negative evidence. **Extract 4.18** below illustrates a positive evaluative recast while **Extract 4.19** shows a re-routing one.

Extract 4.18 Positive evaluative recast

TCH: What are we going to look at today in Citizenship? Ana?
STU: About the %x…x%
TCH: About…excuse me?
STU: About the health
TCH: About being healthy, good.

Extract 4.19 Re-routing recast

STU: eh that the Neolithic people eh.. change of.. of.. no, the Neolithic people live in
TCH: Live ((correcting his pronunciation))
STU: permanently in a.. in a.. in a place and the.. Neolithic or?
TCH: In the Palaeolithic
STU: Palaeolithic. Eh.. in the Palaeolithic they change of.. of place

Within the Communicative Language Teaching framework, Frölich, Spada and Allen (1985) included different types of feedback in their COLT scheme, such as repetition, paraphrase, comment, expansion, and elaboration. For this study, the COLT categories of *expansion* and *elaboration* have been blended into a single one called *expansion*, since the difference between the two was not so clear-cut when analysing the present data. Expansion was the term chosen, as it is a category that appears in many different taxonomies regardless of the paradigm). Similar categories are found in the works of Haneda (2005), Wells (1993) and Sinclair and Coulthard (1975).

In this study the sociocultural distinction between self-repair or other-repair (Van Lier 1988; Aljaafreh & Lantolf 1994) is included in the categories re-route and explicit correction. Other-repair is when it is not the student who corrects a mistake s/he has made, but rather the correction comes from someone external. Self-repair, on the other hand, refers to the student correcting him/herself (see Chapter 3, section 1.1.1). Likewise, the distinction made by many researchers coming from the sociocultural or sociocognitive paradigm between IREs (Initiation-Response-Evaluation) and IRFs (Initiation-Response-Feedback) (Hall & Walsh 2002; Mortimer & Scott 2003; Barnes 1975; Wells 1993; Van Lier 1996) is found to be very useful in this study. As explained in Chapter 3, in IREs the teacher merely evaluates the student response as either correct or incorrect, whereas in IRFs, with the feedback move the teacher boosts the student's learning process beyond the distinction between correct and incorrect responses. IRFs are thought to be more aligned with formative assessment theory than IREs. This is due to the fact

that when teachers evaluate in the third turn, they immediately start a new exchange and no further negotiation is possible; on the other hand, when teachers offer some kind of feedback on the student response, such as expanding on it, students' understanding and learning processes are enhanced. Nassaji and Wells (2000) offer six different possible functions that the third turn may perform: evaluation, justification, counter-argument, clarification, meta-talk, and action. From this taxonomy, I have taken the category *evaluation*, essential in educational talk because of its frequency and also essential for detecting alignment/non-alignment to AfL pedagogy, as mentioned above. This category is also found in the works of Haneda (2005) and Sinclair and Coulthard's (1975) model of classroom discourse analysis.

Also, within the sociocultural and constructivist paradigm, researchers (Alexander 2004; Mercer 2000) claim that the sequence question-answer-feedback needs to be transformed into a productive dialogue in which questions, answers and feedback build coherent and expanding chains of enquiry and understanding (Alexander 2004: 26). Following this approach, and specifically drawing on Wragg and Brown's (2001) framework, I included the categories *answer/comment acknowledged* (teacher nods, smiles or says *yes*, *right*...) within positive evaluation; *repeated verbatim* (teacher re-states students' response and converts it into a question) within re-route-prompt; *praise contribution* (teacher praises contribution, maybe with elaboration) within positive evaluation and expansion; *corrected* (teacher corrects incorrect part of an answer or asks others to correct) within negative evaluation with or without explicit correction; *prompted* (teacher asks prompting questions when first answers are inadequate giving hints to pupils) within re-route-prompt; *probe* (follow-up questions when the first response was inappropriate but requiring more precise and thoughtful answers than prompts) within re-route-prompt.

Regarding the study of feedback from a formative assessment perspective, Tunsall and Gipps (1996) created the categories "specifying attainment/improvement" and "constructing achievement/the way forward". As explained in Chapter 3, these types of feedback are descriptive. The former specifies either successful attainment or how something can be improved. The latter, "constructing achievement/the way forward", focuses on discussing with the student their competence and achievement or rather on discussing with students future possibilities of learning, giving pupils greater responsibilities (Tunsall & Gipps 1996). The category of *meta-feedback* designed for the present study is precisely based on these categories. It refers to the instances in the data in which the teacher offered students a kind of feedback that established success criteria and how pupils' work could be improved to meet those criteria (hence specifying attainment/improvement), and other times in which the teacher's feedback made students reflect upon their learning processes and possible future steps to take (thus

constructing achievement and the way forward). Also working within AfL theory, Ruiz-Primo and Furtak (2006) refer to teachers rephrasing, revoicing or elaborating pupils' answers, which echo some of the final categories used in the present study, specifically positive evaluation and expansion.

The feedback categories applied in this thesis are, then, taken from different research paradigms, namely SLA, sociocultural theory and AfL theory. To sum up, Lyster and Ranta's (1997) categories of *prompts*, *recasts*, and *explicit correction* have been included in this final taxonomy; from the COLT scheme, the category of *expansion/elaboration* has been used as well; within the sociocultural framework, *evaluation* is one of the main categories used in this study; and, finally, one of the most important categories, *meta-feedback*, comes from AfL theory, specifically from Tunsall and Gipps' study (1996).

After referring to the theoretical approaches that have inspired the model designed for the analysis of feedback, I will now explain and exemplify this model in a more detailed way below.

4.2.5.2 Evaluation

Evaluation refers to the teacher assessment of a student's response as correct (positive) or incorrect (negative). Regarding *positive evaluation*, the student answer is assessed as correct through various linguistic means, such as repetition, explicit positive adjective etc. This positive evaluation can include a *recast*, or just the positive remark, with *no recast* incorporated. When the positive evaluation is accompanied by a recast, this does not usually have a corrective function (Lyster 1998b). **Extracts 4.20** and **4.21** illustrate, respectively, *positive evaluation with recasts focusing on content* and *both language and content*. **Extract 4.18** above shows *positive evaluation with recast focusing on language*.

Extract 4.20 Positive evaluation focusing on content

TCH: %x...x% yes, think about the size, think about how much you eat, but what else?
STU: With your family
TCH: With you fam, eating with your family? Yes, sharing, sharing, that's a good one

Extract 4.21 Positive evaluation with recasts focusing on both language and content

TCH: So Pablo Picasso is very famous because he was also?
STU: Abstract
TCH: He was also a painter that did cubism… cubist sorry paintings.

In **Extract 4.20**, the teacher positively evaluates the previous student's response, including a *recast on content*, since what the student says is grammatically correct. In **Extract 4.18**, the *recast* clearly focuses *on language*, as the teacher corrects "the health". In **Extract 4.21**, it seems that the teacher is recasting both content and language, content in the sense of differentiating the artistic concepts of "abstract" and "cubist", and language in the sense of expanding the sentence and be clear about the terminology. These types of recasts (evaluating) have to be distinguished from the type of recast that I have called *re-routing* (see section 4.2.5.4 below), as in this case the learners do not have the chance of uptake after the recast but instead the teacher continues with another initiation. Both within *evaluating* and *re-routing recasts*, it was decided to analyse whether they focused on content or language, whenever possible. This distinction seemed especially interesting given the nature of the classrooms we are dealing with, to see whether there is mainly negotiation of meaning, form, or there is a balance between the two (counterbalanced approach). In the same way, as translation is involved many times in recasts (on language) in these classrooms, the analysis will allow us to see how CLIL is capable of adapting and being flexible, since meaning has a translingual value. That means that meaning goes beyond the code of the message, be it the L1 or the L2. In fact, the L1 may be necessary sometimes.

It may be the case that positive evaluation, as explained above, is not accompanied by a recast (*positive evaluation with no recast* (as illustrated in **Extract 4.22**):

Extract 4.22 Positive evaluation with no recast

TCH: Julio, are you singing again?
STU: No, I'm not!!
TCH: ((teacher makes "taken aback" face)) <u>I'm not singing, perfect English</u>

Moving on to negative evaluation, it refers to the teacher assessing the student's response as incorrect. It may or may not include explicit correction. Explicit correction is included when the teacher overtly states the response is incorrect and offers an explanation on it or the correct form. *Negative evaluation with explicit correction focusing on content* is illustrated in **Extract 4.23**, while **Extract 4.24** exemplifies *negative evaluation with explicit correction focusing on language*, and **Extract 4.25** is an example of *negative evaluation with explicit correction that focuses on both content and language*.

Extract 4.23 Negative evaluation with explicit correction focusing on content

TCH: A <L1 espada L1> swordfish, Alejandro, is not a mammal, it's a fish, ok?
STU: <L1 Porque acaba en L1> fish. Fish. Fish

Extract 4.24 Negative evaluation with explicit correction focusing on language

TCH: So she said the word "is" and he said the word "was". Who's correct?
STU: Was
TCH: Let's read the question. The question. The question is "what's the word that means" ((Gloria is also reading it in a soft voice)), are we in the present tense or past tense?
STU: Past tense
TCH: What'S. What IS the word that means..?
STU: With apostrophe "s"
TCH: The word IS, so this is a per- a perfect, perfect situation of this, eh.. you will be able to write correctly using the grammar. What is the word that means? The word IS. The word was no

Extract 4.25 Negative evaluation with explicit correction focusing on content and language

TCH: Cantabrian coast, north of the Peninsula. Instead of living in r- in rectangular houses, the Celts preferred to build
STU: Square
TCH: Round houses, ok? They were not rectangular, they were round houses, ok?

In **Extract 4.23**, the teacher is correcting the categorization of swordfish as a mammal. In **Extract 4.24**, the *explicit correction* focuses on form because the teacher is focusing on a mistake related to tense, highlighting that students have to use the present tense and not the past. In **Extract 4.25**, illustrating *negative evaluation with explicit correction*, the explicit correction has to do with both content and language. The correction on content refers to the form of Celtic houses. The correction on language has to do with the inclusion of the word "rectangular", hinting at the fact that what the student meant was not square but rectangular.

In negative evaluation with no explicit correction, the correct response is not provided by the teacher (see **Extract 4.26**)

Extract 4.26 Negative evaluation with no explicit correction

TCH: Ok, so people now are following directions. I have blue eyes is wrong. Why is it wrong? Rocío, why is it wrong?
STU: Because we are talking about out perso- personality

As illustrated in **Extracts 4.23-25**, *negative evaluation with explicit correction* both specifies that something was wrong and provides the correct answer, whereas *negative evaluation with no explicit correction* just states the incorrectness of an answer, but the correct alternative is not provided.

4.2.5.3 Expansion

The second type of feedback is *expansion* and it includes two different functions. One refers to the teacher's attempts to push students to continue talking by making further questions with a different function from that of the first question (this would equal the category *elaboration* in the COLT scheme). An example of this is a meta-cognitive question following a question for fact. The second one would be an extension of the content through the addition of related information, but the extension is made by the teacher and not through questions to students (Frölich, Allen & Spada 1985). **Extract 4.27** illustrates this type of feedback:

Extract 4.27 Expansion: further questioning

TCH:	Not eating a lot of meat, ok, why do you think that we should not eat a lot of meat?
STU:	Eh, because, it's bad for the heart ((pronounced incorrectly))
TCH:	Bad for your health, yeah, who can expand on that? Who can tell me a bit more about that? Alejandro?
STU:	For to not have %L1…grasa…L1%

Extract 4.28 Expansion: new information

TCH:	So do you think they were nomads?
STU:	No
TCH:	No. The main difference is that in the Neolithic period the tribes stayed in one place. They didn't move from one place to another.

The teacher, in **Extract 4.27**, is trying to force students into adding more information through the use of further questions. The second function of *expansion* is exemplified in **Extract 4.28**, in which it is the teacher herself who expands the student's answer with new information.

4.2.5.4 Revision

The third type of feedback is called *revision*. Revision is found when the teacher revises and/ or summarizes concepts and content, as instantiated in **Extract 4.29**.

Extract 4.29 Revision

TCH: So you have to write six sentences. And your level is high enough to play with that. If you wrote eh, first I'm going to lift my leg. When I lift my leg, the quadriceps contracts
STU: Contracts
((student is talking simultaneously while teacher explains))
TCH: Ok, and when the quadriceps contract, the opposite muscle, quadri….relaxes. It gets…so you know how to do that right?

In this extract, the teacher and students are revising or summarizing what they have seen during part of the lesson, which is how the muscles in the arm and in the leg work. The teacher is revising it first, because it is a complex notion, and second, because it is an important concept for students to meet the content goals of the unit.

4.2.5.5 Re-route

The next type of feedback is called *re-route*. It is defined as the teacher's attempt to guide students to the right answer after a wrong response. This can be done through different mechanisms, namely prompts and recasts (see below). Although the term corrective feedback in the SLA tradition may match the concept of re-route (Pica 1995; Lyster & Ranta 1997; Gass & Varonis 1994; Sheen 2004; Swain 1995; Long 1996), the term re-route was preferred for two reasons: a) it may include more mechanisms than corrective feedback, especially aiming at reorienting rather than correcting; b) the term corrective feedback is inseparable from the SLA tradition, which is not the main theoretical underpinning in which this study is framed.

Following Lyster and Ranta (1997), *prompts* include elicitations, clarification requests, metalinguistic or metacognitive clues, and repetitions. In addition, and referring back to re-route including more mechanisms than corrective feedback, it also encompasses other strategies such as repetition or reformulation of a question (see **Extract 4.30** below) and a sequence of simpler questions before going back to the original and apparently more complex question (see **Extract 4.32**) (Wragg & Brown 2001). As already mentioned, whenever possible, a differentiation between focus on form or content will be made. **Extract 4.30** illustrates a prompt focusing only on content; **Extract 4.31** is an example of a prompt which focuses on language; and a prompt focusing on both aspects is instantiated in **Extract 4.32**.

Extract 4.30 Prompt focusing only on content

TCH: positive comment about María's sentence
STU: I am funny and intelligent
TCH: But can you tell her something positive about her sentence?
STU: Ah.. I am.. a capital letter and full stop
TCH: A capital letter and full stop.

Extract 4.31 Prompt focusing only on language

TCH: Tell me something about Javier Gonzalez. Tell me something ((raising hand)) participation, participation, participation. Victor, tell me something.
STU: She put the…
TCH: She??
((laughter))
STU: He
TCH: He, he, please respect my Javi!

Extract 4.32 Prompt focusing on both content and language

TCH: It's a crocodile, so?
STU: They are covered
TCH: They belong to which group?
STU: To.. reptiles
TCH: Yes, it is a reptile. But reptiles were divided into?
STU: Four
TCH: Four different groups, so crocodiles belong to which group? ((no answer)) […] Ok, which are the four reptile groups?
STU: Crocodiles, snakes, lizards and turtles
TCH: So the crocodile belongs to..
STU: Crocodiles

In **Extract 4.30**, after an unsuccessful student response, the teacher decides to repeat the previous question, making the student understand their comment was not adequate and leading to an appropriate response in the next contribution. The second example of *prompt* (see **Extract 4.31**) would equal Lyster and Ranta's category of repetition (1997), that is, the teacher repeats part of the student response with a rising intonation in order for the student to realize there is something wrong. The student indeed shows uptake and immediately corrects it. Finally, in **Extract 4.32**, the teacher is asking about the reptile group the crocodile belongs to. Since the student does not seem to know the answer with the first question, the teacher starts asking a series of questions that set the steps the student has to take in her thinking to arrive at the correct response. When all those necessary steps have been taken, the teacher asks the original question again, and now the student responds correctly.

Recasts are defined as reformulations of students' utterances made by teachers, minus the error (Lyster & Ranta 1997). In this study, some recasts are have been considered as *re-route* because they perform a more explicit corrective feedback function rather than a positive evaluative one and because students might have the chance of uptake (whether they take it or not) (see Lyster 1998 on the ambiguity of recasts). As some researchers have proved (Sheen 2006; Sheen & Ellis 2011; Lyster 1998b), the recasts which are more likely to be noticed as negative evidence and be followed by uptake are those which are explicit (as opposed to implicit), which are usually shorter and more emphatic. Below, **Extract 4.19** (already presented above) exemplifies a re-routing recast focusing on content and **Extract 4.33** instantiates a re-routing recast focusing on language.

Extract 4.19 Re-routing recast focusing on content

STU: eh that the Neolithic people eh.. change of.. of.. no, the Neolithic people live in
TCH: Live ((correcting his pronunciation))
STU: permanently in a.. in a.. in a place and the.. Neolithic or?
TCH: In the Palaeolithic
STU: Palaeolithic. Eh.. in the Palaeolithic they change of.. of place

Extract 4.33 Re-routing recast focusing on language

STU: We sing and dance
TCH: We? We sang and?
STU: we sang and [danced]
TCH: [Danced], ok?

In both extracts, the teacher uses her turn to recast the previous student's response. After the recast, the teacher indicates the student that there is an opportunity for uptake, be it through intonation (**Extract 4.33**) or be it through silence. In **Extract 4.19**, the recast focuses on content (Palaeolithic vs Neolithic), whereas the recast in **Extract 4.33** focuses on language, specifically on past tenses.

4.2.5.6 Meta-feedback

The last type of feedback is called *meta-feedback*. This category refers to the feedback that focuses on students' learning and assessment. By saying students' learning, here we mean that this type of feedback focuses on students' weak areas and/or improvements, what they can do to improve learning. By saying students' assessment, this type of feedback may also concentrate on marks and the criteria for reaching each mark, thus being helpful for self- and peer-assessment. An example of this type of feedback is illustrated in **Extract 4.34**.

Extract 4.34 Meta-feedback 1

TCH: What are you learning today?
STU: To write neatly
TCH: Julio's not respecting the teacher
STU: I can't see
TCH: I can't see. One minute, one minute, one minute. I'll ask the question again but first we have to respect <x…x> the class. ((showing the paper for respect)). Most of the time, all the time is a <L1 bien L1> so I hope that you are honest class at the end of the class, because we still have some more minutes in the class.

As illustrated in this extract, the teacher is making reference to the specific criteria students must fulfil to reach a good mark.

Also considered as *meta-feedback* is the positive, negative or neutral evaluation of an activity at a global level, also sometimes including information as to what needs to be done in order to improve it (illustrated in **Extract 4.35**). In other words, when the teacher specifies whether the final product of an activity meets the assessment criteria successfully or not (and why).

Extract 4.35 Meta-feedback 2

STU: That's.. that's ok? ((TCH looks at it and nods))
STU: But Marta, the ear <L1 no me sale L1> ((erases it))
TCH: But they looked ok, Sergio.
STU: Marta? Marta, the mouth.. is bad?
TCH: No, no, no, no.
STU: Yes, yes, yes
TCH: It's fantastic

In these teachers' interventions (**Extract 4.35**), the teacher is letting the student know whether the way in which he is doing the activity (a drawing) fulfils the requirements appropriately, since it seems the student is unsure about how well he is doing the work.

As a final note on **Table 4.7**, in the categories marked with an asterisk a distinction between focus on content or focus on language was made whenever possible. This distinction could not be made in all subcategories. For instance, in the categories of positive evaluation with no recast or negative evaluation with no explicit correction the distinction was not possible because they were utterances of the type "right", "very good", or "that's not correct".

The link between these feedback categories and my research questions will be made clear at this point. The two research questions that are related to feedback are the following:

- What are the interactional features and strategies which characterize AfL discourse in Primary CLIL classrooms?
- Are these patterns specific for AfL classes or are they also found in similar classes where AfL is not implemented?

Next, as the above research questions are general, I will explain and make explicit the expectations that this thesis poses regarding feedback types and types of classes in relation to these two research questions. Regarding *evaluation*, it will be expected that more frequent IRE sequences are found in Non-AfL schools. On the contrary, regarding *expansion* and *meta-feedback*, they are expected to be more frequent in AfL schools. The reason to believe IREs will be less frequent, and *expansion* and *meta-feedback* more numerous in AfL schools is that dialogic interaction is expected to occur more frequently when AfL is being implemented. Within *evaluation*, whenever *negative evaluation* is found, it will be expected to be followed by some kind of *re-routing* in the classes following an AfL pedagogy. In turn, *negative evaluation with explicit correction* will not be expected to be frequent in AfL schools, since AfL is about promoting learning and guiding that learning, and *explicit correction*, by providing students with the correct response, prevents them from reasoning and thinking until they get at it. Moving on to another main feedback type, *revision* is considered an important type of feedback for clarifying concepts and helping students in their learning. In this way, it may be hypothesized it will be more frequent in AfL classrooms because it is aligned with formative assessment theory. Lastly, *re-route* is a category that is more in line with formative assessment than explicit correction, and so it would be expected to be found very frequently in AfL classes. However, the hypothesis is that this is a feature that will also be observed in Non-AfL schools.

4.3 PROBLEMS IN THE CATEGORIZATION

This section will explain what problems were found when applying the methodology and doing the analysis and what decisions were made in order to solve those problems in the best interest for the study. The problems encountered had to do with initiations, responses, and feedback, that is, with the three elements in the IRF pattern. As in any analytical process which involves assigning categories, problems appear, and consistent decisions need to be made in order to ensure the reliability of the analysis. In this case, problems were discussed with other expert researchers in the field to find a solution.

4.3.1 Problems in the categorization of initiations

Within this category, most of the problems had to do with teachers' initiations, although some student initiations also posed difficulties.

In some occasions, certain questions could be ambiguous as whether to be classified as language questions or as content questions (for facts, explanations…). The criterion used to distinguish between these two categories was whether they could be expected to be found in content classes in the L1: if they were, they were classified as content questions; if they were questions that are not expected to occur in the context of a parallel content class in the L1, then they were classified as language questions. The first case is illustrated in **Extract 4.36**; the second one, in **Extract 4.37**.

Extract 4.36 Ambiguous: Content question

TCH: Ok, who can explain what values are?

Extract 4.37 Ambiguous: Language question

TCH: How do you say %L1…avisar…L1%? Maria?

The teacher question in **Extract 4.36** is a question that could be found in any Citizenship class, both in the L1 and in the L2, since it is an important concept for the rationale of the subject. On the contrary, question in **Extract 4.37** could only happen in a class taught through an L2, as the teacher is asking for a translation into that L2.

Some questions contained words such as "explain" that could indicate they were questions for explanations. However, for its categorization, the focus was placed on the genre they triggered, as teachers use the term explain for a number of functions, not just that of explaining. Genres make explicit the types of learning expected in a subject area and the types of language which embody that learning (Veel & Coffin 1996: 194). For instance, a question like the one found in **Extract 4.38** below was classified as a question for fact/definition, even though it says "can you explain", as the teacher is asking for a definition of values. In this context, "Explain what values are" is equivalent to "Define values":

Extract 4.38 Ambiguous: Question for fact/definition

TCH: No? Ok, who can explain what values are?

There are moves that can be both initiation and feedback at the same time, performing both functions, as some researchers have claimed (Sinclair & Coulthard 1975; Van Lier 1988; Wells 1993). In cases like this, in which the same move was performing two different functions, they were double-coded. **Extract 4.39** is an illustration of this:

Extract 4.39 Ambiguous: Initiation + Feedback

TCH: Healthy or not? Ok, for example, healthy, healthy, hands up, healthy or not? ((shows flashcard of man eating pizza)) Mercedes?
STU: No
TCH: Why?

In this extract, the student's response is correct, but the teacher wants to unveil the thinking process that led the student to respond "no". The teacher's question "why?" is, at the same time, a feedback move (expansion) and a new initiation (meta-cognitive question). It was very frequent to find the same utterance with the function of *expansion* as a feedback move and of *meta-cognitive question* as another initiation move.

Those instances in which the teacher gave the floor to a student because s/he wanted them to participate, were not considered initiations, as it is the teacher who willingly gives them the floor. Consequently, the student intervention is not actually an initiation itself (see **Extract 4.40**).

Extract 4.40 Elicited participation

TCH: Yes, it is. Watermelon is natural, it's a fruit, a lot, gives you lots of liquids, ok, and it cools you down. Can you…when it's hot you need to take things that make you cooler, to cool you down. Uh, Pedro??
STU: That I think that boy is very clever because he does not eat ice cream because it's hot and he eats a watermelon that is healthier.

4.3.2. Problems in the categorization of responses

The two main difficulties found when analysing student responses were when students spelt words and when they repeated the same answer in the same turn. In the case of responses involving spelling words (see **Extract 4.41** below), the response could be just one letter or more than one (either completing the whole spelling of the word or not). These instances occurred just in some of the sessions taught by one of the teachers, and they were not easily classifiable under the categories created. It was decided that they be classified as *one-phrase T-unit*. They were not frequent enough as to create a new category of student response and so it was thought

that it fit best under *one-phrase T-unit* than under any of the others. This decision was adopted in agreement with another expert.

Extract 4.41 Response involving spelling words

TCH: Now, can you spell vibrate?
STU: (some)) V-
TCH: Three, two, one, read
STU: ((some)) v-i
STU: ((some)) V-a

When a response was repeated in the same turn by the student (see **Extract 4.42** below), it was not counted twice but one. That means that the response was analysed without taking the repetition into consideration. This was decided because repetitions occurred due to different reasons, such as when the students thought the teacher did not hear them, when the students wanted to emphasize something. In any case, those are not factors that affect complexity of the response, which is what is under scrutiny in the present study.

Extract 4.42 Response repeated within the same turn

TCH: Ok? So a boy is blowing into an instrument, ok? So this is sound. And in the other picture what do you see? Is it a clock or a watch?
STU: ((All)) a clock
TCH: A watch. A clock is on the wall..
STU: A clock, [a clock]

4.3.3. Problems in the categorization of feedback

Complications in this category stemmed from ambiguity and from the fact that the same feedback move could perform various functions at the same time. For example, recasts and explicit correction could be confused in some contexts, as in **Extract 4.43** below:

Extract 4.43 Recast or explicit correction?

TCH: She wants to hug the boy, and how do you know that? What is the tree doing with its branches? Ruben
STU: He is going to…
TCH: She is…
STU: She is going to…to hug the boy.

Teacher's feedback in instances such as that illustrated above was classified as *recasts* (within *re-route*). The reasons for this classification were that explicit correction is normally

accompanied by a negative evaluation (the teacher explicitly indicates that what the student has said is incorrect) and chance for uptake on the student's part is not usually given, for the right version is given by the teacher him/herself (Lyster & Ranta 1997; Lyster 1998a; Sheen 2004; Sheen & Ellis 2011). In contrast, with this type of *recast* (within *re-route*) students are given the opportunity of uptaking the wrong utterance. This is explained by the fact that some recasts are more explicit than others, as it is the case of the ones within *re-route* (Lyster & Ranta 1997; Lyster 1998a; Ellis & Sheen 2006; Sheen 2006; Sheen & Ellis 2011; Llinares & Lyster 2014).

Finally, there were cases in which a feedback move could have two different functions at the same time. For instance, feedback moves that were classified as both *expansion* and *revision* (illustrated in **Extract 4.44**). This took place in revision episodes, where most of the feedback offered by the teacher could be considered *revision* but sometimes it could also be considered *expansion* if we just considered the previous student's response.

Extract 4.44 Expansion + Revision

TCH: So they walk on four.. how are eh.. lizards' limbs, are they big or are they small?
STU: Small
TCH: They are small, no? They have small limbs. So in Spanish the word is <L1 reptar L1>, ok? But the word ah.. crawl we use it only for.. to.. to explain how lizards move from one place to the other. Same we use slither when we wanna talk about how snakes move from one place to another, ok? So to crawl is the way they.. they move... with four limbs and those four limbs are very, very short, so that's how they move.

This section has, then, summarized and explained the problems and solutions found at the time of data analysis, going through teachers' initiations, students' responses and teachers' feedback. As already made evident, at certain times, the analysis scheme needs to be flexible rather than fixed, as in this way it will give us more possibilities and fewer limitations.

4.4 SUMMARY AND CONCLUSION

This fourth chapter has been devoted to explaining the corpus collected for this research and the methodology used for the analysis of the data, as well as to describing the problems encountered when coding the data, and the decisions taken to overcome those problems. In the data section, the characteristics of the corpus have been described: information on the schools, the teachers, the students, the number of sessions, and hours recorded. In the same way, the methodology of the collection and transcription of the data has also been explained. The second section of the chapter has addressed the methodology of the analysis and has been divided into five different

parts: episodes (the first layer of analysis), teachers' initiations, student responses, students' initiations and teachers' feedback. In each of them, the categorizations and taxonomies used have been presented and illustrated, and their theoretical bases have been explained. Finally, the last section has been concerned with explaining and illustrating the different problems that arose when data analysis was carried out and the corresponding solutions that were adopted. As a final remark, let us underline that this chapter is crucial for understanding the following ones on results and discussion.

5

Results I — Episodes

This is the first in a series of chapters which present the results of the analyses of classroom interaction and AfL in CLIL primary classrooms in the Community of Madrid. As a reminder, the data analysed consist of 47 lessons of approximately 50 minutes each, collected in 4 different schools (two of them were implementing AfL, the other two were not). In each school, the same teacher was recorded teaching two different subjects, and in each subject two different didactic units were recorded. The present chapter focuses on *episodes*, which are the different parts or stages in a lesson in which teacher and students are carrying out a specific activity (Snell & Lefstein 2011). In this chapter, first, there will be a comparison of the episode types and frequencies for each type of school (AfL vs. Non-AfL). Then, the different episodes will be illustrated and described: those common to both types of contexts, those only found in AfL classes, and those only found in Non-AfL classrooms. Next, the types of episodes will be analysed in relation to the type of subject. Finally, the different episodes will be explored for each of the four teachers individually.

As a reminder, not all the episodes were analysed as the focus was on interaction and not all the episodes involved meaningful interaction between teacher and students or among students (see Chapter 4). By meaningful interaction, what is meant is interaction about content, language, or both. If students did not participate in the discourse, or they merely read aloud, these episodes or parts of episodes were not taken into account for analysis. The episodes that were not analysed were *classroom management*, *reading from the book*, *listening from the book*, *dictation*, and *singing songs*. Other episodes, such as *stating objectives for the lesson*, *individual activity*, *group/pair activity*, *explanation of homework/activity*, and *explaining marks* were only partially analysed. Although these episodes were partially analysed or not analysed in terms of discourse (questions, responses, feedback), they are still present in the quantitative part of the episode analysis.

5.1 TYPES OF EPISODES IN AFL AND NON-AFL CONTEXTS: QUANTITATIVE ANALYSIS

This section is devoted to presenting the different types of episodes that appear in each type of school. **Table 5.1** and **Figure 5.1** illustrate the different types of episodes found in the AfL schools (taught by Teacher 1 and Teacher 2) and in the Non-AfL schools (taught by Teacher 3 and Teacher 4). Focusing first on similarities, *whole-class discussion* and *class management* episodes are the most frequent types of episodes in both AfL and Non-AfL classes, whereas *introducing the topic* seems to be a very infrequent one in both types of schools. *Explanation of homework/activity* is another episode that is regularly present in both schools. Moving on to differences between school types, AfL schools have more episodes (double the amount in Non-AfL schools) and a slightly greater variety as well per total number of words. It is interesting that the only episodes which are not found in the AfL schools are *listening from the book,* and *dictation*, perhaps as they entail little active participation from students. In Non-AfL, non-present episodes are *stating objectives for the lesson*, *explaining marks*, and *singing songs*. In AfL schools, within the least frequent episodes, we find *reading from the book*, *singing songs*, and *introduction of topic*. In Non-AfL schools, the least frequent episodes were *students presenting their work*, *dictation*, and *self-/peer-assessment*. Other interesting differences are that there is more *revision of lesson* in the case of Non-AfL schools, and that *individual activities* have more weight than *group/pair activities* in this context, in contrast with the AfL groups.

	AfL		Non-AfL	
Type of episode	N	%	N	%
Whole-class discussion	98	23.7%	67	29.1%
Classroom management/class routines	94	22.8%	53	23.0%
Explanation of activity/homework	46	11.1%	33	14.3%
Activity: group/pair discussion/work	47	11.4%	10	4.3%
Self- and peer- assessment	43	10.4%	3	1.30%
Activity: students presenting their work	15	3.6%	3	1.30%
Students doing activity individually	11	2.7%	15	6.5%
Revision of lesson (last and current)	9	2.2%	19	8.3%
Evaluation/Correction of activity/homework	8	2.0%	9	4.0%
Reading from the book	4	1.0%	8	3.5%
Introduction of topic	5	1.2%	4	1.7%
Stating objectives for the lesson	18	4.4%	—	—
Explaining marks	12	2.9%	—	—
Singing (songs)	3	0.7%	—	—
Listening from the book	—	—	4	1.7%
Dictation	—	—	2	0.9%
Total number of episodes	413		230	

Table 5.1 Episodes in AfL schools and Non-AfL schools: frequency of occurrences

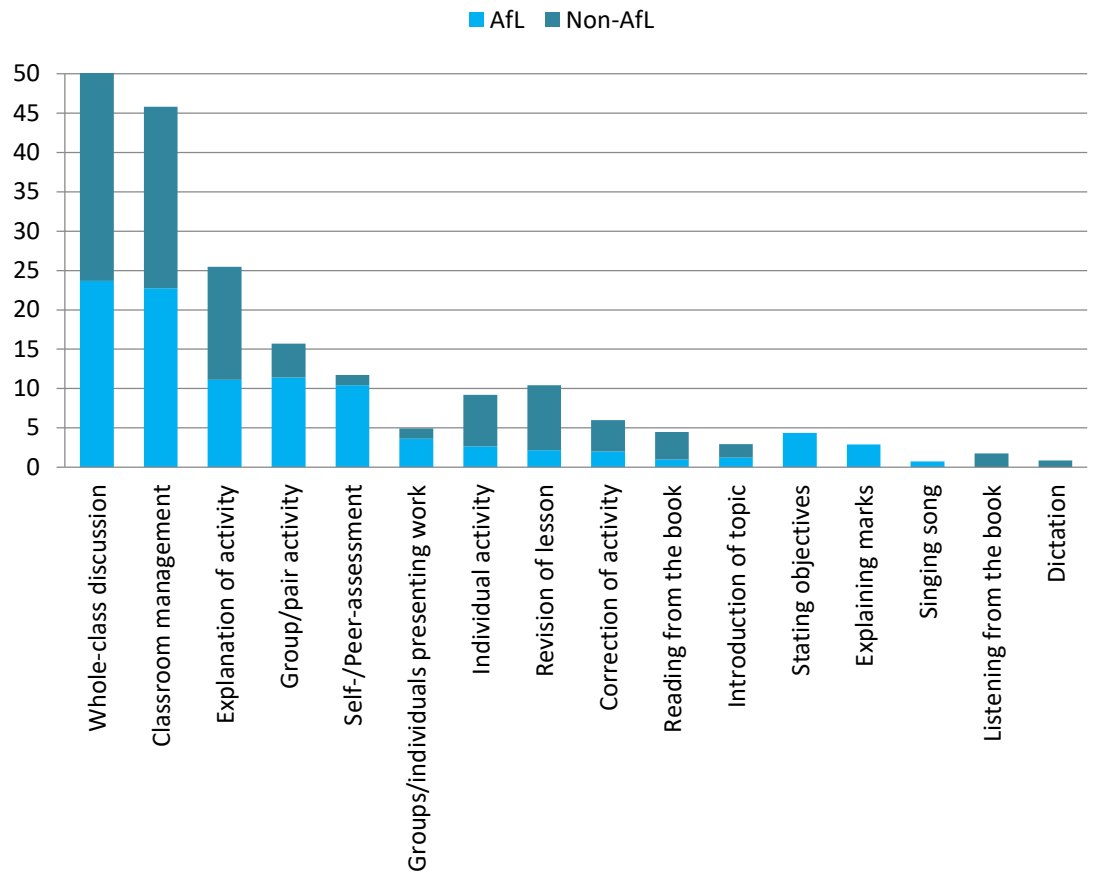


Figure 5.1 Episodes in AfL schools and Non-AfL schools.

Regarding types of episodes, then, in spite of a few commonalities, differences between both types of schools are quite notable. A more qualitative account of these commonalities and differences is presented in the next section.

5.2 TYPES OF EPISODES IN AFL AND NON-AFL CONTEXTS: QUALITATIVE ANALYSIS

This section will qualitatively illustrate some of the episodes identified by providing extracts taken from the corpus. It will be divided into three subsections: episodes which appear in both AfL and Non-AfL classrooms, episodes which are mainly present in AfL classes, and episodes that generally appear in Non-AfL classrooms.

5.2.1 Frequent episodes in both AfL and Non-AfL classrooms

The two most frequent episodes in both types of schools are *whole-class discussion* (23.7% in AfL schools, 29.1% in Non-AfL schools) and *classroom management* (22.8% and 23% in AfL schools and Non-AfL schools, respectively). In **Extract 5.1** we have an example of a *whole-class discussion* episode, in which teacher and students are discussing mammals, specifically the characteristics all mammals share.

Extract 5.1 Whole-class discussion (Teacher 3, Non-AfL, Science, first unit, class 1)

TCH:	No? Ok, so the majority of you brought mammals, so let's look at mammals, at all the mammals that are on the board 'cause we're gonna start studying mammals. Can you tell me something that you see all of them have in common? Look at the.. No, think about it. Look at all the different mammals. We have a walrus, a hamster, an elephant, a mole, a rhino, a horse, a dog, panda bear, a polar.. I think this is not a polar bear.
STU:	Yes!
TCH:	This is a white fox, but anyway.
STU:	No, is a polar bear
TCH:	It's a fox. It is a.. a white fox
STU:	Can I look?
TCH:	A lynx, a tiger, cheetah and a hamster, so think about it. One thing do they all have in common.
STU:	are vertebrates
TCH:	Wait a second.
STU:	They are [mammals]
TCH:	[Yes, but] no, no, no, no, no. Wait a second. They are all vertebrates, we know that
STU:	And mammals
TCH:	And we know
STU1:	they %x think x%
TCH:	You have to raise your hand, Rodrigo. And they are all mammals, but I wanna see if you can think of other things they have in common. <L1 A ver L1>, Natalia
STU2:	they all have four legs
TCH:	Good, so they have four legs.
STU:	The walrus have legs?
STU:	no, the.. the walrus no
STU:	No!
STU:	and human
TCH:	Ok, so they have.. we will talk about it later, so that means that they have limbs, no? Our limbs are our hands and our.. our arms and our legs, no? Those are our limbs, so mammals have also limbs, so that could be one characteristic they all have in common.

In **Extract 5.1**, the teacher is discussing with the whole class the characteristics of mammals. Instead of her reciting all of them, she tries to get students to think about those characteristics by looking at the pictures of mammals. Students start with the obvious ("they are vertebrates" and "they are mammals"), but the teacher encourages them to think beyond those evident characteristics.

Extract 5.2 illustrates an episode of *classroom management*, in which students are getting ready, following the teacher's instructions.

Extract 5.2 Classroom management (Teacher 1, AfL, Science, second unit, class 1)

TCH: … but let's start with the text, because the time is going very, very quickly and we have to start looking at this incredible text from.. razkids, ok? I have the plastics.. please, take them out of the plastics and give me the plastics, ok? Very quickly. What can you say? ((she is giving out the texts))
STU: ((All)) can you give me one?
TCH: Can you give me one? Can I borrow? Can I have this interesting black and white <x…x> text?
STU: <x…x> do we have to <x…x>?
TCH: Give me that plastic
STU: Therese, and me
TCH: ohhh! Ok, quickly go to your desks. Be sure that your homework is out and <x…x> with the <L1 bien L1>.

To finish illustrating the episodes that are common to both school types, **Extract 5.3** shows the episode *explanation of activity/homework*, in which Teacher 2 is explaining the learners the activity about slogans they are going to do in groups.

Extract 5.3 Explanation of activity/homework (Teacher 2, AfL, Citizenship, second unit, classes 1 and 2)

TCH: What I want you to do is... OK, we're going to... about two weeks ago, we looked at adverts and slogans, do you remember doing that?
STU: Yes
TCH: Yeah?
STU: Yes
TCH: And you were all really good at thinking of a slogan that was eye catching and will be able to sell your product, yes?
STU: Yes
TCH: Right. And we need World Health Day, OK? We need to think about telling other people about being healthy. OK?, so I'm going to put you in groups, and each group would have a different part of being healthy and you have all to think of a slogan, to design, to design a poster, to %x...x% around the school about being healthy. One group will be advertising healthy food. Eat healthy food and what could be the slogan? "Eat good, feel good". Yeah, OK, think, things like that. One group would be doing food; one group would be doing exercise; sleep, we all need enough sleep, if we don't have enough sleep, like Itziar was saying she becomes angry, get in a really bad mood...; relaxation, you need to make your mind really switch off, OK?, you need to do what feels good for you; and hygiene, washing our hands, brushing our teeth, having a shower, changing of shorts, that %x...x% that you had for a whole week, that's horrible! OK? Hygiene is really important as well. In the groups you... are with just now, OK?

In the case of **Extract 5.3**, the teacher is explaining an activity which students are going to do in class and in groups. She first contextualizes the activity with reference to a previous activity from a previous unit, and then she explains the task students have to do. To boost students' confidence, in the contextualization, the teacher emphasizes how good students were at a previous similar activity.

After seeing some of the common episodes appearing in both school types, the next subsections will illustrate those episodes that seem to be more characteristic of AfL schools and those which seem to be more frequent in Non-AfL schools.

5.2.2 Episodes mainly present in AfL schools

There are two episodes which only appeared in the AfL data analysed. These are *explaining marks* and *stating objectives* (illustrated in **Extract 5.4** and **Extract 5.5**, respectively). Likewise, *self- or peer-assessment* was barely present in Non-AfL schools, as opposed to AfL schools (see **Extract 5.6**). The three of them are very closely related to the implementation of AfL (hence their absence in Non-AfL schools), in which it is essential that students know how to assess themselves and their classmates, and they know what is expected of them and what the assessment criteria are.

Extract 5.4 Explaining marks (Teacher 1, Science, first unit, class 1)

TCH: You need…. we need to know if you have a Sufi, a Bien, Notable, a Sobre or what. So here around the room, there is a poster ((points)), there is a poster ((points)) there is a poster ((pointing)) and here is a poster, ok? So, for the difficult things a Sufi is a Sufi. ((reading aloud)) I can recognize some phenomena produced…. produced…. produced by sound. I can describe the pitch and the loudness. These words I don't understand! Is that ok?

STU: Yes

TCH: Yes, because we just started, so during the unit, you will have to understand these things for a Sufi. You will have to understand more things, because you have to explain some phenomenon. Explain. That's more difficult. But we have two days to do it. And here in the Notables: ((reading)) explain how sound moves through solids and liquids and gasses.

In **Extract 5.4**, the teacher is explaining students what they have to do to reach a "sufi" (equivalent to a C) and a "notable" (equivalent to a B+). She does not only explain what students have to do to get certain marks, but she also has put posters on the walls with that information, so that students can read them and look at them whenever they need to throughout the unit.

Extract 5.5 Stating objectives for the lesson (Teacher 2, Citizenship, second unit, first class)

TCH: I'm going to show you WALT and WILF like this. %x…x% isn't it? I don't have to keep writing it. Can you see that? Is that big enough or not?
STU: Yes
TCH: Yes? Ok, who would like to read the WALT for me? Lucas? Could you please?
STU1: ((reading)) Understand what being healthy means. That there are many things we need to think about to be healthy. What these are and why they are important. Understand the reason why we celebrate World Health Day. Know how to tell people effectively ((pronounced incorrectly)) about health.
[…]
TCH: WILF, who'd like to read the WILF out? María? Would you %x…x%?
STU2: ((reading WILF)) I can tell you about my healthy ((should be "health")) and what being healthy means. I can look at somebody else…else…he…health ((pronounced incorrectly))
TCH: Health ((correction))
STU2: Health and anal…analy…analyse ((pronounced incorrectly))
TCH: Analyse ((correction))
STU2: Analyse it.

When stating the objectives for the lesson, *WALT* and *WILF* are two important elements of AfL. They represent, respectively what *We Are Learning Today* and *What I'm Looking For*. The former refers specifically to learning objectives whereas the latter is more related to success criteria. In **Extract 5.5**, the teacher is sharing both aspects with the students, so that it is clear what is expected of them. WALT and WILF are not only shared and discussed with students, they are also accessible to students at any time, since they are posted on the walls.

Extract 5.6 Self-assessment (Teacher 2, Citizenship, first unit, second class)

TCH: Ok we're going to have to stop there, Adrian, ok, ((reading from the WALT and WILF)) to talk about different types of emotions, you've talked about your own emotions in one day, talking about how you felt, what…why you felt that way, ok, and what you did afterward, ok? Do you think that looking at the WALT and the WILF, read through those quickly, %x…x% you need to look at them to read them, Mateo they are over there, and can you show me if you think that we have accomplished what we wanted to learn today. ((students showing thumbs up/ down)) Victor what do you think about today? Do you think we've learned more about how to deal with what happens? Yes, no, or so-so….. no, you don't, ok. Aitana yes? Can I have everybody, please? Daniel, you're not showing me how you feel…thank you. This table so-so, Sandra?
STU: Yes
STU: So-so, ok, yes? Hugo? Alejandro?
STU: Yes
STU: Yes

In **Extract 5.6**, the teacher, who is encouraging self-assessment, reminds students that they must do it by taking into account WALT and WILF. These two AfL elements help students

understand the learning objectives they have to reach in each lesson and in each unit. Having those goals clear, the task of self- or peer-assessment is facilitated.

It is worth noting that these three types of episodes seem to promote the types of questions (meta-questions) and the type of feedback (meta-feedback) that students need to reflect on their learning and assessment processes and to check their learning and progress against the objectives of the lesson or unit (see Chapter 6 for types of questions and Chapter 8 for types of feedback).

Another episode that has a heavier presence in AfL schools than in Non-AfL ones is that of *group/pair discussion/work*. In **Extract 5.7**, students are doing pair work, practicing how to act out a conversation they have previously written down, while the teacher goes around the class helping them if they have problems.

Extract 5.7 Pair /group activity/work (Teacher 1, Citizenship, first unit, class 3)

STU:	Hello, my name is Álvaro!!!
	((students practicing, making a lot of noise in the class))
TCH:	Energy and gestures!
STU:	How are you?
STU:	I'm not a very %x…x% and I am tall and my height is one meter and %x…x% and I am funny because I like% x…x% and you?
STU:	Fine thank you.
STU:	Today I'm going to %x…x% look at this!
STU:	Can you %x…x%?
TCH:	Everybody should be practicing if you're sitting you're not practicing very well.
	((some students stand))
STU:	I am the first %x…x%
STU:	Very tired!
STU:	No, I'm not.
STU:	The last day
TCH:	Ah, the last day. You have to write a story. We only have seven minutes, ooh!!

This subsection has exemplified those episodes only or largely appearing in AfL schools. The next one will do the same in the case of Non-AfL schools.

5.2.3 Episodes mainly present in Non-AfL schools

This subsection will illustrate those episodes that only appear in Non-AfL schools, as it is the case of *listening from the book* and *dictation*, and some other episodes that are more frequent in Non-AfL schools than in AfL ones, such as *revision of lesson* and *individual activity*. **Extract 5.8** and **Extract 5.9** are illustrations of the first group: *dictation*, *listening*, and *reading from the book* (*reading from the book* also appears in AfL schools, although it only represents 1%).

Reading from the book is exemplified along with *listening from the book* because one normally follows the other. In **Extract 5.8**, students are listening to some information they have written on the book and they later read that same information.

Extract 5.8 Listening and reading from the book (Teacher 2, Science, first unit, class 2)

TCH Now we are going to.. ((students are talking and she has to speak really loud)) we are finished.. we have finished with mammals and we are going to read, if we can
STU: Birds
TCH: about birds, ok? We haven't finished, the classroom is not over, open your book. ((students are making noise)). We are going to listen to the information about birds.
[…]
TCH: Ok, so let's listen to the information ((She plays the CD)) Ok. Let's start reading here. Laura, you start ((Laura starts reading a few sentences, and then other students read when the teacher orders them so)).

It seems that *listening* and *reading from the book* do not give teachers the opportunity of asking students questions to make them think and expand their knowledge. In **Extract 5.9** students have to write what the teacher tells them in the way of a dictation

Extract 5.9 Dictation (Teacher 4, Citizenship, second unit, class 3)

TCH: It's a dictation. Shhh. It doesn't matter, %x...x% OK? Good. So... ((starts dictating)) Generalising
STU: No, no, no, no...
TCH: Shhh. Generalising: from general, generalising.
STU: Is... %L1... separado?...L1%
TCH: One word. Generalising, generalising about... Gabriel, we are %x...x% and...
STU: Ah!
TCH: %x...x% I'll repeat. Generalising about particular groups... OK, I repeat: Generalising about particular groups of people. I repeat: Generalising about particular groups of people and... this is a difficult word, OK?, I'm going to spell it for you: labelling, OK?, label, like the label of the clothes ((Points at his t-shirt)) So l-a-b-e-l-l-i-n-g. I repeat, labelling, l-a-b-e-l-l-i-n-g. I repeat: Generalising about particular groups of people and labelling them... labelling them. Double l, i-n-g, them
STU: Did you say then or them?
TCH: Themmm. OK, I repeat: Generalising about particular groups of people and labelling them "comma", we are
STU: we are?
TCH: We are creating, we are creating false... We are creating. Paula, %x...x% We are creating false expectations, false expectations that some individual, that some individual members, that some individual members of the group have certain, "certain", I write it on the blackboard, certain, do you understand certain?
STU: No
STU: OK, we'll explain this later. OK. Certain negative traits, certain negative traits, traits, do you understand?
STU: %L1...Yes, tratos...L1%
STU: %L1...Tratos...L1%

TCH: Mmm characteristics, OK?, traits. Traits ((writes it on the blackboard)) t-r-a-i-t-s, traits, shhh. Certain traits, well certain negative traits, OK?, that have been attributed, with double t, attributed ((Someone sneezes)) Bless you. Attributed to the group, OK?
STU: To?
TCH: To the group, attributed to the group, OK?, yes.

Some of the episodes that are more frequently present in Non-AfL than in AfL schools are illustrated in **Extract 5.10** and **Extract 5.11**. In **Extract 5.10**, teacher and students are *revising* last lesson together (due to its length, just part of the episode is included in the extract); in **Extract 5.11**, students are *doing an activity individually* for the Arts class, specifically building a boat.

Extract 5.10 Revision of lesson (Teacher 3, Science, first unit, class 2)

TCH: We started the new unit called vertebrates, ok? And we were studying the general characteristics of?
STU: ((some)) [mammals]
STU: ((some)) [vertebrates]
TCH: Mammals, we started with mammals. So very quickly ok? I asked you to write down in your notebook the different characteristics all mammals have, yes or no?
STU: ((all)) yes
TCH: So we are going to review them very quickly. Take out your notebook and if you don't have.. if you don't have one characteristic, then complete your chart, ok? So we are going to.. Alejandra! Alejandra and Milena, ok? We are going to check all the characteristics that you wrote in your chart, alright? So, very quickly, we are going to review them.
[…]
TCH: Ok. Let's go ((writing on the blackboard)). Mammals. Jessica
STU1: mammals are viviparous
TCH: Mammals are viviparous. Very good. They are viviparous ((while writing it on the board)). Who can explain me what does it mean? That mammals are viviparous. Alicia
STU2: they born eh.. from the- their mother bo- body
TCH: They are born from their mother's?
STU: ((Alicia and some other students)) Body
TCH: Ok. Another characteristic. María Belén
STU3: eh...
[…]
STU3: mammals live in different habitats.. habitats. Eh.. are.. they are terrestrial or... they are some terrestrial or aquatic
TCH: Ok, so mammals live in many different habitats. Some of them live on land, some of them live in the air and some of them live?
STU: In [water]
TCH: [In water]
STU: Marta
TCH: Who can tell me what is the only mammal that lives.. or that can fly? The only mammal that can fly. Yolanda
STU4: the bat
TCH: Bat

This *revision* episode triggers a series of IRF sequences. In episodes in which students work individually on an activity (**Extract 5.11** below), the teacher goes around the class to check students' work and solve their doubts. In the case of **Extract 5.11**, the individual activity students have to do in the Arts class is building a boat.

Extract 5.11 Individual activity (Teacher 4, Arts, first unit, class 1)

TCH: Some people didn't finish the boat, remember that we were building a boat, a %x….x%. You didn't finish, you were %x…x%
STU: You haven't builded the boat
TCH: All right,
((some students get up to go and retrieve their boats))
((students whispering and working quietly, teacher getting organized))

This subsection closes the qualitative analysis of episodes in the two types of schools (AfL and Non-AfL), those which are common to both schools as well as those which are more frequent in one or the other. Next, the different episodes across subjects will be presented.

5.3 TYPES OF EPISODES ACROSS SUBJECTS

This part of the chapter will analyse the types of episodes in each of the four subjects analysed, without taking into account the type of school and the teacher. Likewise, since each of the teachers was recorded teaching two different subjects, the different episodes across the two subjects will also be examined in this section. The purpose is to discover whether the type of subject affects the types of episodes appearing most or least frequently.

5.3.1 Types of episodes across subjects

This subsection shows the different episodes appearing in each of the four subjects present in the corpus: *Science*, *Citizenship*, *Arts*, and *Drama* (see **Table 5.2** and **Figure 5.2**). No differences between the two didactic units recorded in each of the different subjects were found.

Episode/Subject	Science	Citizenship	Arts	Drama
Classroom management/class routines	47 (22.0%)	73 (21.8%)	18 (34.0%)	9 (23.1%)
Whole-class discussion	66 (30.8%)	79 (23.6%)	11 (20.7%)	7 (17.9%)
Explanation of activity/homework	17 (8.0%)	50 (15.0%)	6 (11.3%)	6 (15.4%)
Activity: group/pair discussion/work	16 (7.5%)	33 (9.8%)	1 (1.8%)	7 (17.9%)
Self- and peer- assessment	15 (7.0%)	27 (8.0%)	1 (1.8%)	3 (7.7%)
Activity: students presenting work	—	12 (3.5%)	2 (3.7%)	4 (10.3%)
Revision of lesson	12 (5.6%)	12 (3.5%)	3 (5.6%)	1 (2.6%)
Students doing activity individually	7 (3.2%)	10 (3.0%)	9 (17.0%)	—
Evaluation/Correction of activity/homework	9 (4.2%)	7 (2.1%)	1 (1.8%)	—
Reading from the book	6 (2.8%)	6 (1.8%)	—	—
Stating objectives for the lesson	4 (1.8%)	12 (3.5%)	—	2 (5.1%)
Explaining marks	5 (2.3%)	7 (2.1%)	—	—
Introduction of topic	4 (1.8%)	4 (1.2%)	1 (1.8%)	—
Singing song(s)	2 (1.0%)	1 (0.3%)	—	—
Listening from the book	4 (1.8%)	—	—	—
Dictation	—	2 (0.6%)	—	—
Total number of episodes	214	335	53	39

Table 5.2 Episodes across subjects: frequency of occurrence (in relative descending order)

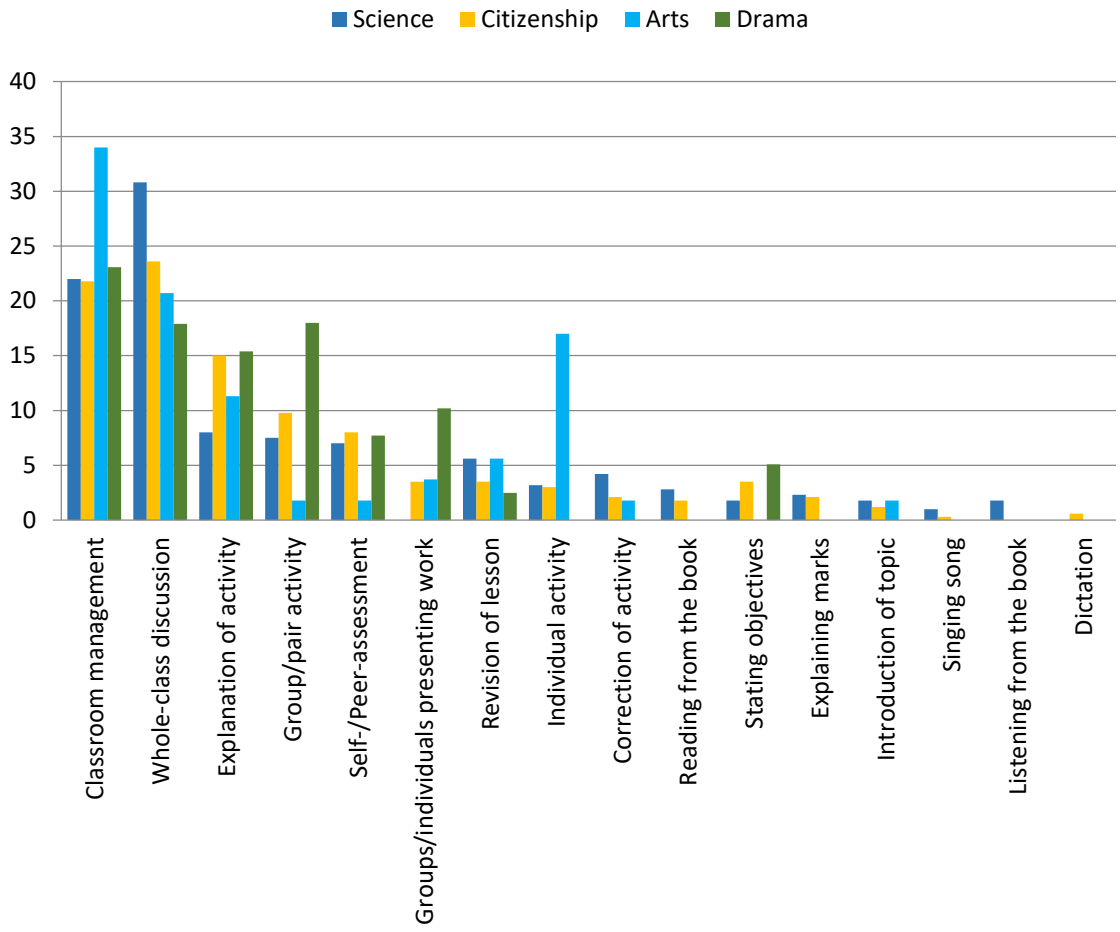


Figure 5.2 Episodes across subjects.

In all four subjects, the most frequent type of episode is *classroom management*: 22% in Science, 21.8% in Citizenship, 34% in Arts, and 23.1% in Drama. Likewise, *whole-class discussion* episodes are very frequent, too, although less frequent in Drama lessons (28% in Science, 21.8% in Citizenship, and 20.7% in Arts, as opposed to 17.9% in Drama). The episode *explanation of activity* is relatively frequent in the four subjects, too (8% in Science, 15% in Citizenship, 11.3% in Arts, and 15.4% in Drama). Among the least frequent across subjects, we find *revision of lesson, stating the objectives for the lesson, reading from the book, explaining marks, introduction of topic, singing songs, listening from the book, dictation*. This is due to the fact that some of these episodes mainly or exclusively occur in one of the two types of schools (see section 5.2.). Apart from *whole-class discussion, classroom management*, and *explanation of activities*, the three subjects, Science, Citizenship and Drama, are characterized by more group or pair work than individual students working on their own, as opposed to Arts, in which individual activities predominate. In the case of Drama, not only *group/pair work* is frequent, but also *individual activities* and *students presenting their work to the rest of the class*. Other findings contrasting Arts with the other three subjects are that episodes of *self- or peer-assessment* are very infrequent, and the episode *stating objectives* is not present. In general, both Science and Citizenship have a greater number and variety of episodes than Arts and Drama, which seems to be independent from the type of school.

5.3.2 Same teacher teaching two different subjects: are there any differences regarding episode types?

This subsection will be concerned with differences and similarities regarding types of episodes found in the lessons of each teacher teaching two different subjects. At this point, it may be worth repeating the topic of each didactic unit (see **Table 5.3** below).

	Subjects	Didactic unit
Teacher 1	Science	Sound
		Bones and Muscles
	Citizenship	A circle of smiles
		The Giving Tree
Teacher 2	Drama	Word association
		Improvization
	Citizenship	Emotions
		Being healthy
Teacher 3	Science	Vertebrates
		Pre-History and Pre-Roman times
	Arts	Cubism
		Pop art
Teacher 4	Citizenship	Democracy
		Gender
	Arts	The Alhambra
		Parallel lines

Table 5.3 Didactic units

As illustrated in **Table 5.3** and **Table 5.4**, Teacher 1 taught Science and Citizenship, Teacher 2 Citizenship and Drama, Teacher 3 Science and Arts, and Teacher 4 Citizenship and Arts. Since the variable *teacher* is controlled here, with this analysis we can confirm whether differences in the episodes can be really attributed to the type of subject. As in the previous section, significant differences between the two didactic units recorded in each of the subjects were not found.

	Teacher 1		Teacher 2		Teacher 3			Teacher 4
	Science	*Citizenship*	*Citizenship*	*Drama*	*Science*	*Arts*	*Citizenship*	*Arts*
Whole-class explanation/discussion	44 (30.8%)	36 (19.6%)	11 (23.4%)	7 (17.9%)	24 (32.9%)	2 (10.0%)	32 (30.7%)	9 (27.3%)
Classroom management/class routines	35 (24.5%)	46 (25.0%)	4 (8.5%)	9 (23.1%)	12 (16.4%)	8 (40.0%)	23 (22.1%)	10 (30.3%)
Explanation of activity/homework	10 (7.0%)	21 (11.4%)	9 (19.1%)	6 (15.4%)	7 (9.6%)	2 (10.0%)	20 (19.2%)	4 (12.1%)
Activity: group/pair discussion/work	15 (10.5%)	19 (10.3%)	6 (12.8%)	7 (17.9%)	1 (1.4%)	—	8 (7.7%)	1 (3.0%)
Self- and peer- assessment	15 (10.5%)	22 (12.0%)	3 (6.4%)	3 (7.7%)	—	1 (5.0%)	2 (1.9%)	—
Activity: students presenting their work	—	6 (3.3%)	5 (10.6%)	4 (10.3%)	—	—	1 (1%)	2 (6.1%)
Students doing activity individually	3 (2.1%)	7 (3.8%)	1 (2.1%)	—	4 (5.5%)	3 (15.0%)	2 (2%)	6 (18.2%)
Revision of lesson (last and current)	2 (1.4%)	4 (2.2%)	2 (4.3%)	1 (2.6%)	10 (13.7%)	3 (15.0%)	6 (5.8%)	—
Evaluation/Correction of activity/homework	6 (4.2%)	1 (0.5%)	1 (2.1%)	—	3 (4.1%)	—	5 (4.8%)	1 (3.0%)
Reading from the book	—	3 (1.6%)	1 (2.1%)	—	6 (8.2%)	—	2 (1.9%)	—
Introduction of topic	2 (1.4%)	2 (1.1%)	1 (2.1%)	—	2 (2.7%)	1 (5.0%)	1 (1%)	—
Stating objectives for the lesson	4 (2.8%)	9 (4.9%)	3 (6.4%)	2 (5.1%)	—	—	—	—
Explaining marks	5 (3.5%)	7 (3.8%)	—	—	—	—	—	—
Singing song(s)	2 (1.4%)	1 (0.5%)	—	—	—	—	—	—
Listening from the book	—	—	—	—	4 (5.5%)	—	—	—
Dictation	—	—	—	—	—	—	2 (1.9%)	—
Total number of episodes	143	184	47	39	73	20	104	33

Table 5.4 Episodes in each teacher and their two subjects: frequency of occurrence (in relative descending order).

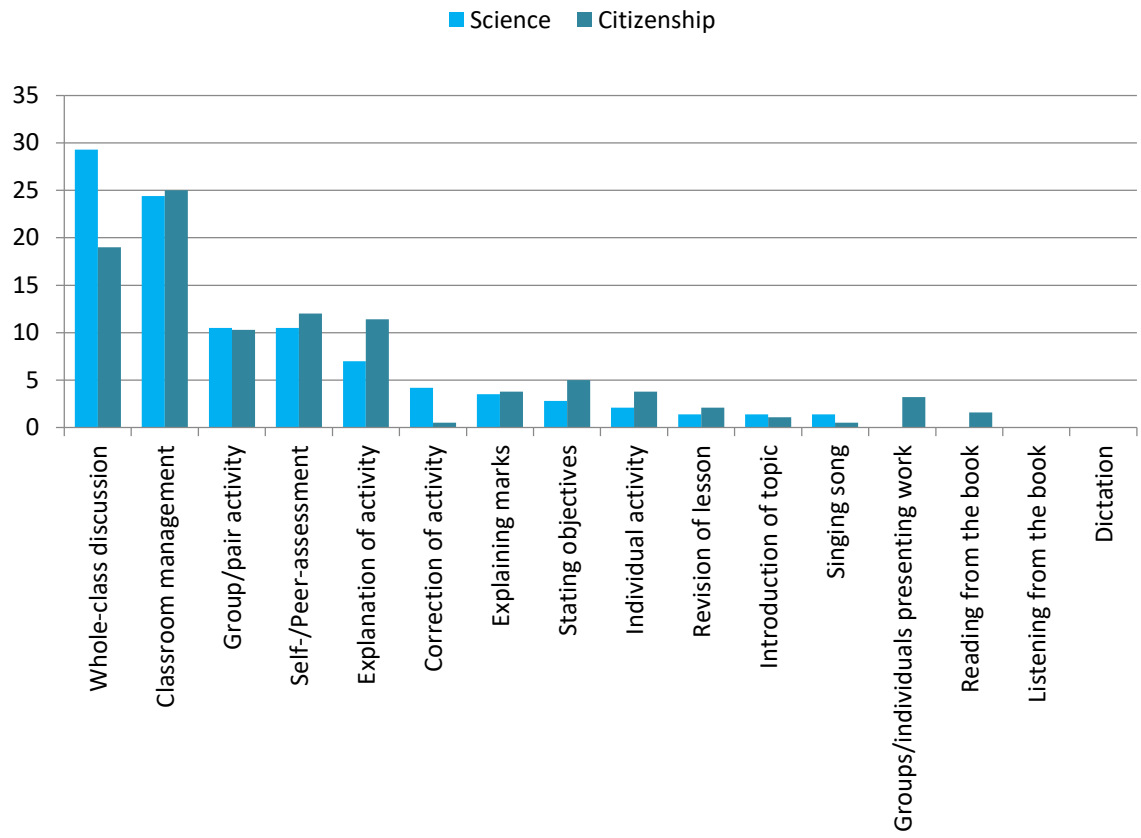


Figure 5.3 Episodes in Teacher 1's Science and Citizenship lessons.

In **Table 5.4** and **Figure 5.3** above, the different episodes found in Teacher 1's lessons are displayed. A lot of similarities can be identified. Among the similarities, in both subjects, *whole-class discussion* and *classroom management* episodes are the most frequent ones. However, *whole-class discussion* is much more frequent in the case of Science. Other similar frequencies are found in the following episodes: *group/pair work*, *explanation of activity*, *self- and peer-assessment*, *revision of lesson*, *stating objectives* and *explaining marks*. The main differences are found in the episodes *evaluation/correction of activity* (more frequent in Science), *students presenting their work to the class* and *reading from the book* (both only found in Citizenship). It seems, then, that no matter what subject Teacher 1 is teaching, certain types of episodes will appear, such as *group/pair work*, *self- and peer-assessment*, *explaining marks*, and *stating objectives for the lesson*. When it comes to these types of episodes, the role of the teacher seems to have more weight than the role of the subject. In other types of episodes, such as *whole-class discussion*, *evaluation of activity*, and *students presenting their work to the class*, it is the subject that appears to play its part.

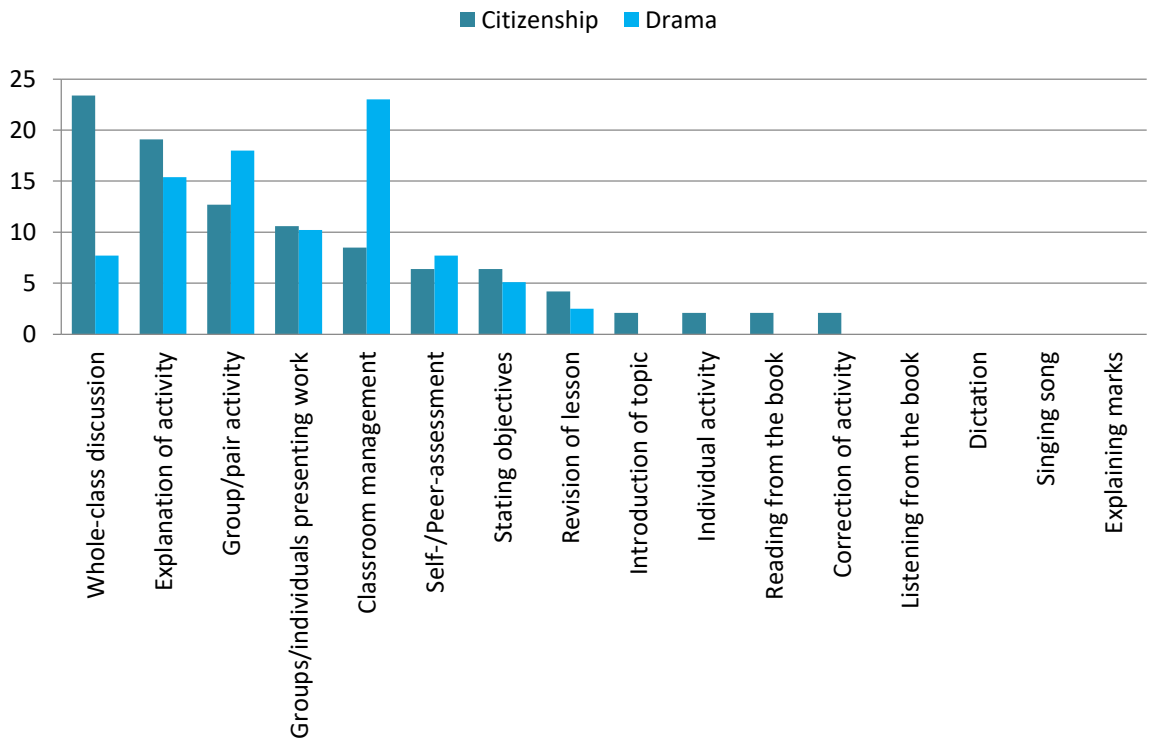


Figure 5.4 Episodes in Teacher 2's Citizenship and Drama lessons.

In Teacher 2's lessons, both similarities and differences can be noticed (see **Table 5.4** and **Figure 5.4** above). There are a number of episodes that score a similar percentage both in the Drama and Citizenship lessons. These are: *revision of lesson*, *stating objectives*, *groups or individual students presenting their work to their classmates*, and *self- and peer-assessment*. Regarding episodes which are more frequent in Citizenship lessons, we can find *whole-class discussion* and *explanation of activity or homework*, whereas the episodes which are more frequent in Drama lessons are *classroom management* and *group or pair work*. Also, there are episodes that appear in Citizenship and not in Drama, as Citizenship lessons tend to have more variety of episodes. Among them, there are *introduction of topic*, *individual activities*, *reading from the book*, and *correction of an activity or homework*. The only episode appearing in Drama and not present in Citizenship is *whole-class activity*. As it occurred with Teacher 1, both the teacher and the subject seem to have an effect on the episode type. For instance, Teacher 2 tends to include episodes such as *revision of lesson, stating objectives, students presenting their work* and *self- and peer-assessment* no matter what subject she is teaching. On the other hand, the variable *subject* seems to have an effect on other types of episodes, such as Citizenship, which displays more *discussion*, and Drama, which contains more *group/pair work*.

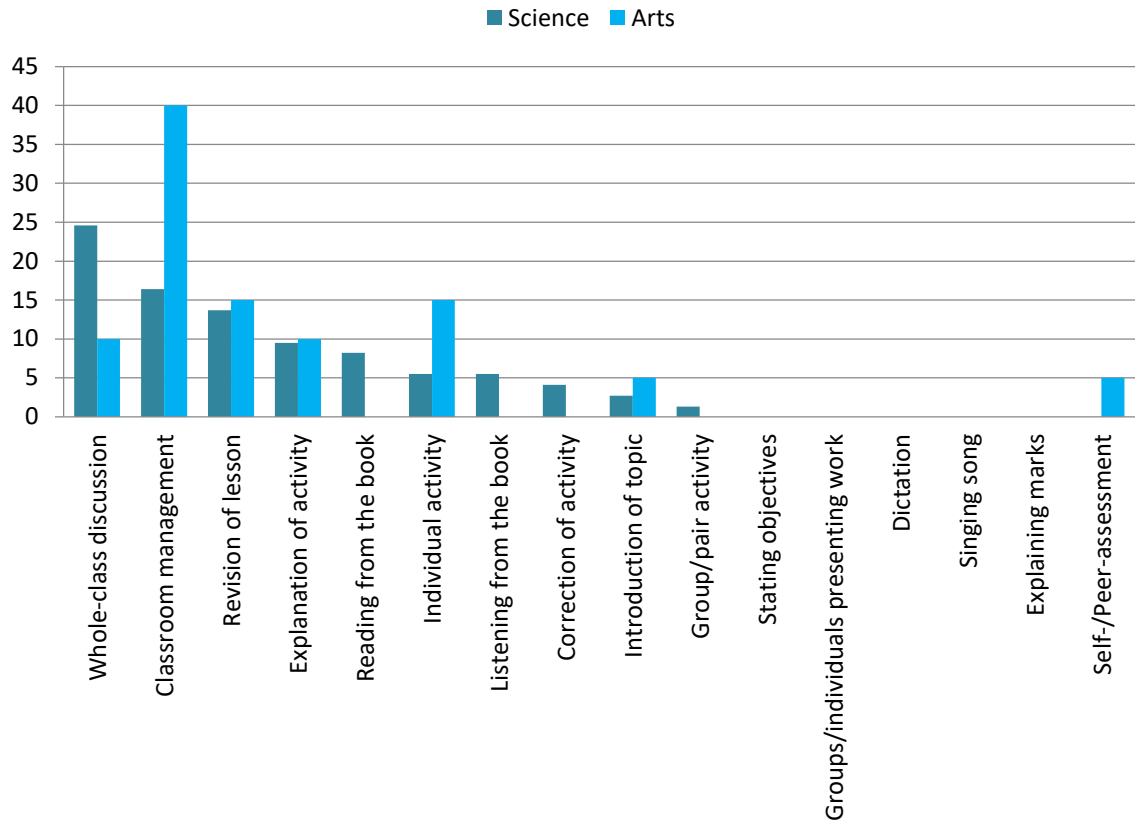


Figure 5.5 Episodes in Teacher 3's Science and Arts lessons.

The episodes in Teacher 3's Science lessons seem to be quite different from those appearing in her Arts lessons (see **Table 5.4** and **Figure 5.5**). Starting with the commonalities, similar percentages are found in *revision of lesson*, *introduction of topic*, and *explanation of activity/homework*. Also, there a number of episodes that do not occur in either of the two subjects: *stating objectives*, *dictation*, *singing songs*, and *explaining marks*. Moving on to the differences, *whole-class activity*, *group/pair work*, *listening and reading from the book*, and *correction of activity or homework* were only found in Science lessons. In turn, the only episode only appearing in Arts and not in Science was *self-/peer-assessment*. Science, then, displays a more varied array of episodes. Regarding the episodes of *classroom management* and *individual activities*, both are much more frequent in Arts, whereas *whole-class discussion* appears more often in Science lessons. In Teacher 3's case, the type of subject apparently plays a bigger role than the teacher herself, as few commonalities can be found between the types of episodes found in her Science lessons and the ones found in her Arts lessons. In this way, Science triggers more *whole-class discussion*, *group/pair work*, and *correction of activity/homework*, whilst Arts seems to prompt more *classroom management* and *individual activities*.

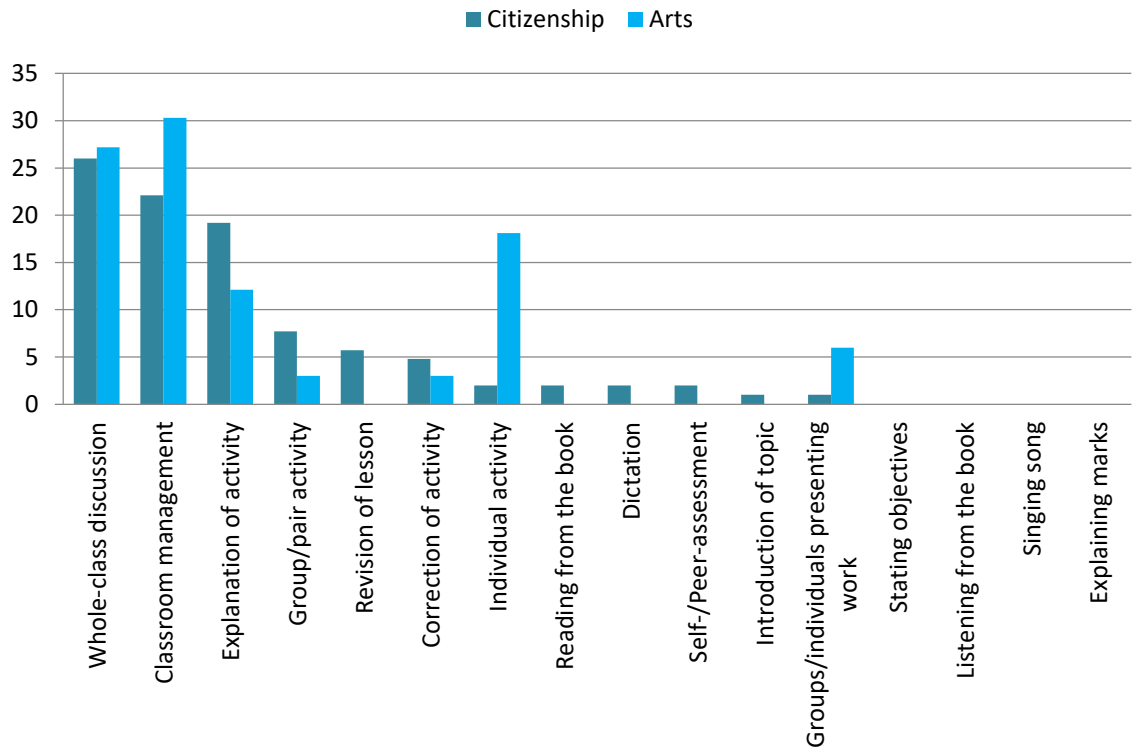


Figure 5.6 Episodes in Teacher 4's Citizenship and Arts lessons.

Finally, **Table 5.4** and **Figure 5.6** show the episodes present in the lessons taught by Teacher 4 in the subjects of Citizenship and Arts. *Whole-class discussion* and *class management* episodes are quite frequent. However, *classroom management* episodes are even more numerous in Arts. *Correction of homework* presents very similar percentages in both subjects. The episodes of *explanation of activity* and *group/pair work* appear more often in Citizenship lessons, whereas *individual activities* and *students presenting their work* to their classmates are more common in Arts. Several episodes appear in Citizenship lessons but not in Arts, such as *revision of lesson*, *introduction of topic*, *whole-class activity*, *reading from the book*, *dictation*, and *self- and peer-assessment*. However, the contrary does not occur, that is, no episodes present in Arts are absent in Citizenship. Therefore, the variety of episodes is greater in Citizenship lessons than in Arts lessons. Other episodes are not present in either Citizenship or in Arts, like *stating objectives*, *listening from the book*, *singing songs*, and *explaining marks*. As it was the case for Teacher 3, the types of episodes found in Teacher 4's lessons seem to depend more on the subject than on

his teaching style: Citizenship produces a wider variety of episodes and it fosters group work, as opposed to Arts, which triggers more individual activities.

In general and focusing exclusively on the type of subject, few differences can be noticed between Science and Citizenship, except for *whole-class discussion* being more frequent in Science and *explanation of activity* and *students presenting their work* being more repeated in Citizenship lessons. Overall, Science and Citizenship lessons are characterized by a lot of *whole-class discussion*, *pair/group work*, *explanation of activity/homework*, and *self-/peer-assessment*. Arts is mainly portrayed as a subject in which a lot of individual work goes on, contrasting with the rest of the subjects. Drama is differentiated from the rest of the subjects in that there is much less *whole-class discussion*, and much more *group/pair activities* and groups of *students presenting their work*. As in the case of the other three subjects, in Drama *classroom management* and *explanation of activity* episodes are very recurrent. For all the teachers, there was a relation between the episodes used and the type of subject, especially in the case of non-AfL Teachers 3 and 4. For Teachers 1 and 2 (AfL), their methodological approach also seemed to be influential.

5.4 TYPES OF EPISODES USED BY DIFFERENT TEACHERS

In this section, the raw numbers and percentages of the types of episodes appearing in all the lessons taught by each of the teachers are presented, without differentiating across subjects. In the case of Teacher 1's lessons, AfL teacher, the episodes most frequently found are *classroom management* and *whole-class explanation/discussion* (see **Table 5.5** and **Figure 5.7**). Similarly, *self-/peer-assessment* and activities in pairs or groups are very frequent. *Reading from the book* and *singing songs* are the episodes with the lowest frequencies. *Listening from the book* and *dictation* are the only episodes which are not found in these classes, thus there is a wide variety of episodes.

Type of episode	Number and %
Classroom management/class routines	81 (24.8%)
Whole-class explanation/discussion	77 (24.5%)
Self- and peer-assessment	37 (11.3%)
Activity: group/pair discussion/work	34 (10.4%)
Explanation of activity/homework	31 (9.5%)
Stating objectives for the lesson	13 (4.0%)
Explaining marks	12 (3.7%)
Students doing activity individually	10 (3.1%)
Evaluation/Correction of activity/homework	7 (2.1%)
Activity: students presenting their work	6 (1.8%)
Revision of lesson (last and current)	6 (1.8%)
Introduction of topic	4 (1.2%)
Reading from the book	3 (0.9%)
Singing song(s)	3 (0.9%)
Listening from the book	—
Dictation	—
Total number of episodes	327

Table 5.5 Episodes in Teacher 1's lessons

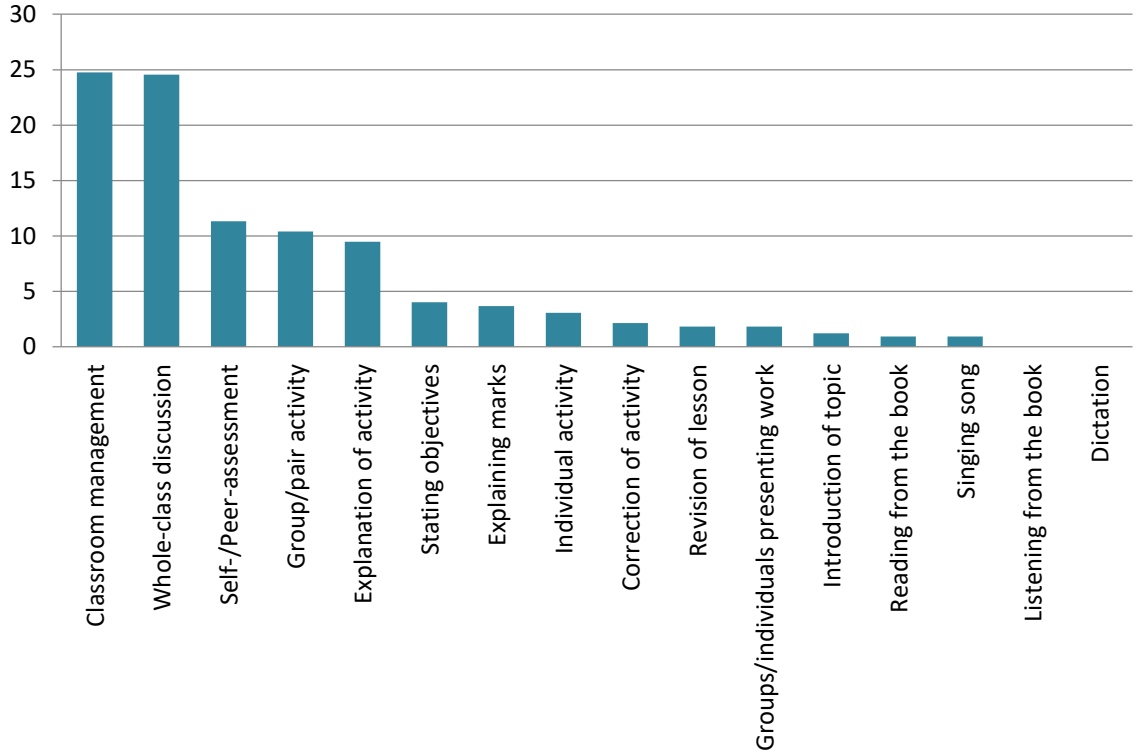


Figure 5.7 Episodes in Teacher 1's lessons.

Both **Table 5.6** and **Figure 5.8** present the different episodes in Teacher 2's lessons, another AfL teacher. There are four types of episodes which do not appear: *listening from the book*, *dictation* (as in Teacher 1's lessons), *singing songs*, and *explaining marks*. *Students doing an activity individually*, *reading from the book*, and *evaluation/correction of homework* or activities are barely present. *Explanation of activity/homework* and *whole-class discussion/explanation* are the most frequent types of episodes (the former due to the high percentage that this episode scored in Drama). *Class management* episodes and *group/pair activities* are also quite frequent.

Type of episode	Number and %
Whole-class explanation/discussion	18 (20.9%)
Explanation of activity/homework	15 (17.4%)
Classroom management/class routines	13 (15.1%)
Activity: group/pair discussion/work	13 (15.1%)
Activity: students presenting their work	9 (10.5%)
Self- and peer-assessment	6 (7.0%)
Stating objectives for the lesson	5 (5.8%)
Revision of lesson (last and current)	3 (3.5%)
Introduction of topic	1 (1.2%)
Students doing activity individually	1 (1.2%)
Reading from the book	1 (1.2%)
Evaluation/Correction of activity/homework	1 (1.2%)
Listening from the book	—
Explaining marks	—
Singing song(s)	—
Dictation	—
Total number of episodes	86

Table 5.6 Episodes in Teacher 2's lessons.

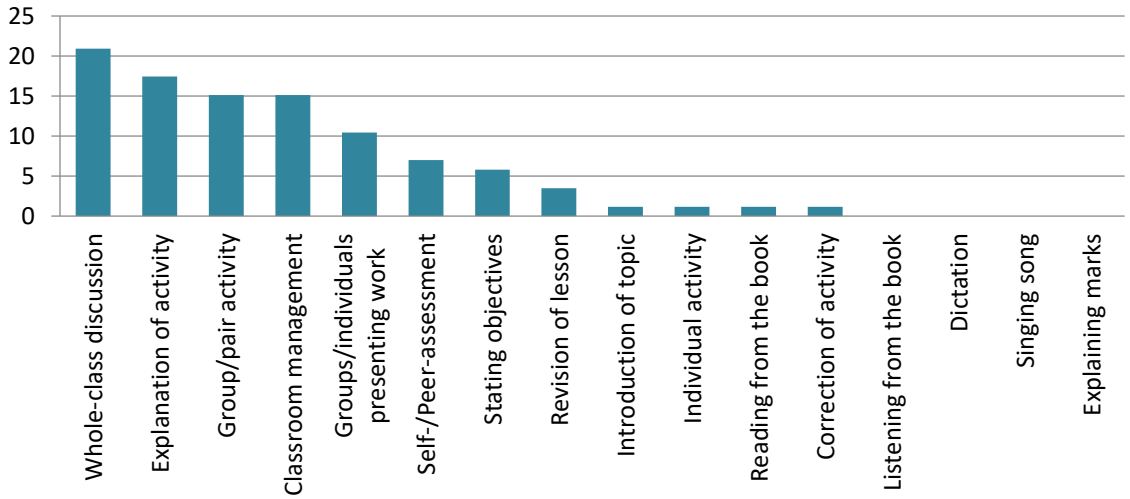


Figure 5.8 Episodes in Teacher 2's lessons.

Table 5.7 and **Figure 5.9** show the types and frequencies of episodes in Teacher 3's lessons, a Non-AfL teacher. There is a lot of variety, though slightly smaller than in the case of lessons taught by Teachers 1 and 2, as there are five types of episodes completely absent, the most remarkable *stating objectives* and *explaining marks*. *Classroom management* episodes and *whole-class discussion/explanation* are the most frequent ones (as with Teacher 1's lessons). *Group/pair activities* and *self-/peer-assessment* are the episodes appearing less frequently (contrasting with lessons taught by Teachers 1 and 2).

Type of episode	Number and %
Whole-class explanation/discussion	26 (28.0%)
Classroom management/class routines	20 (21.5%)
Revision of lesson (last and current)	13 (14.0%)
Explanation of activity/homework	9 (9.7%)
Students doing activity individually	7 (7.5%)
Reading from the book	6 (6.5%)
Listening from the book	4 (4.3%)
Evaluation/Correction of activity/homework	3 (3.2%)
Introduction of topic	3 (3.2%)
Activity: group/pair discussion/work	1 (1.1%)
Self- and peer-assessment	1 (1.1%)
Activity: students presenting their work	—
Stating objectives for the lesson	—
Explaining marks	—
Singing song(s)	—
Dictation	—
Total number of episodes	93

Table 5.7 Episodes in Teacher 3's lessons.

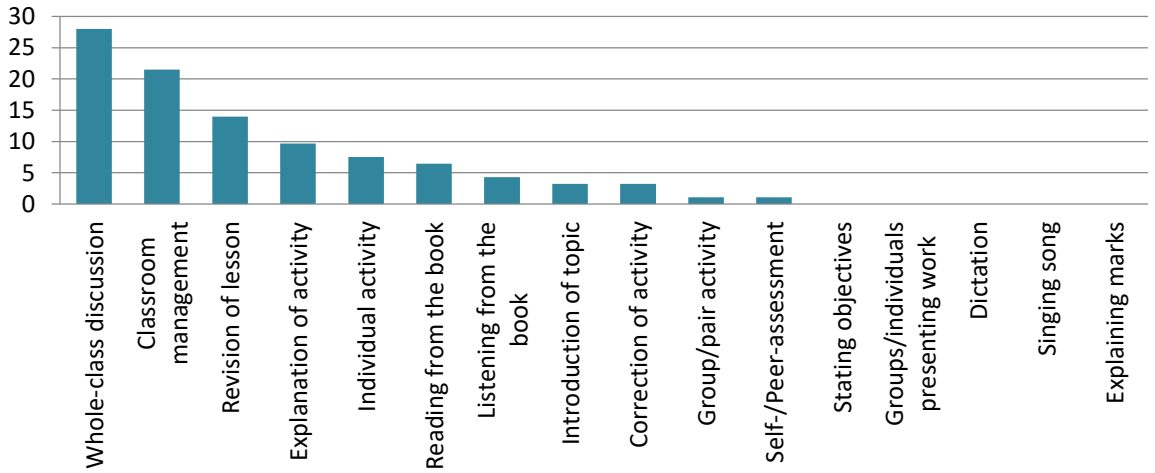


Figure 5.9 Episodes in Teacher 3's lessons.

In **Table 5.8**, the episodes appearing in Teacher 4's (Non-AfL) lessons are illustrated. Four types of episodes are not present, among which we find *stating objectives* and *explaining marks* (in line with Teacher 3's lessons and contrasting with Teachers 1's and 2's). As with Teachers 1's and 3's lessons, the majority of episodes are *whole-class discussion/explanation* and *classroom management or routines*. The least frequent is *introduction of topic*, followed by *dictation*, *reading from the book* and *self-/peer-assessment*, which again aligns more with Teacher 3's lessons and contrasts with Teachers 1's and 2's.

Type of episode	Number and %
Whole-class explanation/discussion	41 (29.9%)
Classroom management/class routines	33 (24.1%)
Explanation of activity/homework	24 (17.5%)
Activity: group/pair discussion/work	9 (6.6%)
Students doing activity individually	8 (5.8%)
Evaluation/Correction of activity/homework	6 (4.4%)
Revision of lesson (last and current)	6 (4.4%)
Activity: students presenting their work	3 (2.2%)
Self- and peer-assessment	2 (1.5%)
Reading from the book	2 (1.5%)
Dictation	2 (1.5%)
Introduction of topic	1 (0.7%)
Stating objectives for the lesson	—
Listening from the book	—
Singing song(s)	—
Explaining marks	—
Total number of episodes	137

Table 5.8 Episodes in Teacher 4's lessons.

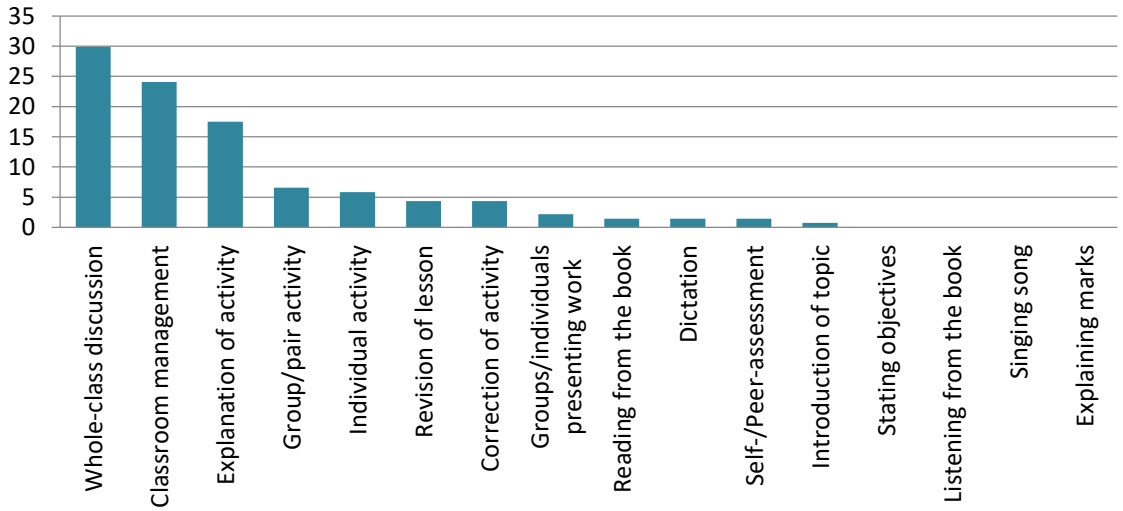


Figure 5.10 Episodes in Teacher 4's lessons.

In general, the types of episodes appearing in lessons taught by Teacher 1 and Teacher 2 are similar, such as the presence of *stating objectives for the lesson* and *self/peer-assessment*. These episodes are not found in the case of Non-AfL teachers. Similarly, *group or pair work* is frequent in both Teachers 1 and 2 and not so frequent in the case of Teachers 3 and 4, whose lessons tend to include more individual activities. *Evaluation/correction of homework* is also more frequent in the case of Non-AfL teachers, as well as *reading from the book, dictation*, and *listening from the book*. Commonalities among the four teachers are also found, such as *whole-class discussion* and *classroom management* episodes being the most frequent ones, and *revision* and *explanation of activity/homework* having a similar presence in all cases (except for Teacher 3 and her use of revision, which is higher than in the rest of the cases).

5.5 DISCUSSION

This section will discuss the results presented in the previous sections, starting with the similarities found between the two types of contexts (AfL and Non-AfL) and continuing with differences. Common to both types of schools is the high frequency of *whole-class discussion* (22% in AfL schools and 24.3% in Non-AfL schools) and *classroom management* episodes (22.8% in AfL and 23% in Non-AfL schools). Many researchers have previously claimed that whole-class discussions are the norm in most classrooms (Lyster 2007: 87; Lyster & Mori 2006; Dalton-Puffer 2006; Fazio & Lyster 1998; Hiebert 1999; Alexander 2004). Nassaji and Wells (2000) even argue that this type of episode is fundamental to develop an inquiry-oriented approach in the classroom.

It seems to be the case that certain types of episodes appear more frequently in AfL schools, in my view, to a large extent, due to AfL methodology. These episodes are *stating objectives for the lesson* (4.4% in AfL schools, non-existent in Non-AfL), *explaining marks* (2.9% in AfL schools, non-present in Non-AfL schools), and *self-/peer-assessment* (10.4% in AfL schools, 1.3% in Non-AfL). Sharing learning criteria with students is a key feature of AfL (Black & Wiliam 1998 a, b; Black et al. 2003; Black et al. 2004; Hattie & Timperley 2007), mainly because if students know what the teacher expects from them, they can improve their performance. At the same time, sharing criteria with learners is crucial if self- or peer-assessment are to be practiced (Sadler 1989). Torrance and Pryor (2001) show how important it is to clarify learning goals and criteria with students through interaction. In the same way, self- and peer-assessment is one of the most important characteristics of AfL (Black & Wiliam 1998 a, b; Assessment Reform Group 1999, 2002; Rea-Dickins 2001). Self- and peer-assessment is essential to learning because they make students agents of their learning (Black et al. 2003;

Black et al. 2004; Leung 2004; Harrison & Howard 2009). Both self- and peer-assessment also help students develop their metacognitive skills (Harrison & Howard 2009; Heritage 2010; Black et al. 2004; Sadler 1989) and are evidence of students' understanding for the teacher, who can then prepare adequate interventions (Heritage 2010; Black et al. 2004).

Thus, there are certain episodes that seem to be more aligned with an AfL pedagogy. Expectedly, these episodes are very frequent in AfL schools and not so much in Non-AfL schools. These are activities in groups or pairs as opposed to the students working individually or other episodes such as reading, listening, or dictating (more frequent in Non-AfL schools), in which students are more passive participants. The results of the comparison of episodes in AfL and Non-AfL schools have shown that, in line with AfL pedagogy, the episodes in AfL schools engage learners and make them active participants in the process of learning, as opposed to more transmissional ways of teaching (Barnes 1975; Black & Wiliam a, b; Leung & Mohan in press; Rea-Dickins 2006; Wiliam et al. 2004; Mohan et al. 2010).

It is also interesting to observe how subject types seem to have an effect on the types of episodes that appear in the lessons. This has been noticed in the case of Drama (in an AfL school) and Arts (in both Non-AfL schools), in which we saw how the number of episodes involving activities (*group/pair discussion/work* and *students doing activity individually*) was higher than in the other two subjects (Citizenship and Science). This may be explained by the fact the Arts and Drama are more *hands-on* subjects, where students are most of the time involved in art projects (such as drawing a cubist painting) or Drama activities (such as improvization). Likewise, it seems that individual activities are more promoted in Arts, as opposed to Drama, where all the activities are carried out in pairs or groups. As for Science and Citizenship, there is more space for whole-class discussions, which create an inquiry-oriented environment, and activities in pairs or groups are preferred over individual ones, thus helping students develop their abilities of cooperative learning.

As for whether it is the subject or the teaching style that has a major effect on the types of episodes, results are not conclusive. In the case of AfL teachers, both aspects seem to play their role: there are common episodes in both subjects, although differences are also found as a result of the subject they are teaching. On the contrary, in the case of Non-AfL teachers, the type of subject appears to have greater weight on the type of episodes, as few commonalities are found. This may be the result of AfL training, since the episodes repeated in both subjects in AfL classes are *stating objectives for the lesson*, *self- and peer-assessment* and *explaining marks*. All these episodes are directly related to the implementation of AfL. Thus, teachers trained in this pedagogy are aware of the importance of these types of episodes if learning is to be fostered and

students are to gain autonomy in their learning processes. All in all, however, the type of subject could be said to have more influence on the types of episodes than the teaching style, since it does have an effect in the case of the four teachers (lesser in the case of AfL teachers, greater in the case of Non-AfL ones).

5.6 SUMMARY

In this chapter, I have presented the different types of episodes that can be found in the data analysed. First, the episodes appearing in each type of school (AfL and Non-AfL) have been presented in a quantitative way. Secondly, the similarities and differences found regarding episode types between the two types of schools have been qualitatively illustrated through extracts from the corpus and have been related to types of questions and feedback in a qualitative manner too. Thirdly, quantitative results of episodes across subjects have been shown, along with the results of the episodes appearing in the lessons of each teacher teaching two different subjects. Finally, the episodes appearing in the lessons of each teacher have been presented, without distinguishing between subjects. In general, it can be concluded that whole-class discussion and classroom management episodes are the most frequent ones and that the subject can determine the specific weight of certain types of episodes (such as individual and group or pair activities being more frequent in Drama or art than in Science or Citizenship). Specifically, comparing both types of schools, in AfL schools there are more and a wider variety of episodes. Also, it seems that teachers in AfL schools devote more time to group work as opposed to Non-AfL schools, in which individual work has more presence. Likewise, the episodes which one would expect to be intimately related with the implementation of AfL (*stating the objectives for the lesson*, *explaining marks* and *self-/peer-assessment*) seem to happen only or to a greater extent in AfL schools. As for individual teachers, *classroom management* and *whole-class discussion* are the most frequent episodes for all teachers. Apart from them, in the case of Teacher 1, *group/pair discussion/work*, *self/peer-assessment*, and *explanation of activity/homework* are also quite frequent. *Dictation* and *listening from the book* are non-existent, and *introduction of topic*, *students presenting their work*, *revision of lesson*, *reading from the book*, and *singing songs* are infrequent. As for Teacher 2's lessons, *students presenting their work*, *group/pair discussion/work*, *explanation of activity/homework*, and *self/peer-assessment* frequently appear. On the other hand, *introduction of topic*, *individual activities*, *reading from the book*, and *evaluation/correction of homework* barely appear. For Teacher 3, individual activities, *revision of lesson*, *reading from the book*, and *explanation of activity/homework* increase. The least frequent episodes are *self/peer-assessment* and *group/pair*

discussion/work. Some other episodes are not present in Teacher 3's lessons: *stating objectives for the lesson*, *explaining marks*, *dictation*, *listening from the book*, and *singing songs*. Finally, Teacher 4, apart from the two most common episodes for all teachers, also uses *explanation of activity/homework* quite frequently. One step behind, there would be *revision of lesson*, *students doing activity individually*, *group/pair discussion/work*, and *evaluation/correction of activity/homework*. Among infrequent episodes, we find *self- and peer-assessment*, *dictation*, *reading from the book*, *students presenting their work*, and *introduction of topic*. *Explaining marks*, *stating objectives for the lesson*, *singing songs*, *listening from the book*, and *dictation* are completely absent from Teacher 4's lessons.

6

Results II — Question types

This is the first chapter devoted to the analysis of the IRF patterns (Sinclair & Coulthard 1975) identified in the corpus under study, specifically to its first part: initiations (I). The chapter specifically focuses on teacher initiations through questions using Dalton-Puffer's (2007) typology (see Chapter 4). This typology divides questions into *questions for facts* (asking for objective happenings), *questions for explanations* (asking for how something happened, elaboration of facts), *questions for reasons* (asking for causes why something happened), *questions for opinions* (asking for students' personal opinion about an issue), and *meta-cognitive questions* (asking students to argue a viewpoint or articulate their thinking). In addition to this taxonomy, two other categories were added: *language questions* (further sub-divided into *content-obligatory*, which ask for the language required to master the content, and *content-related*, which relate to other linguistic aspects) and *meta-questions* (those which make students reflect about their learning/assessment, and those which make students assess a classmate's work).

This chapter is divided into different sections. First, frequencies and comparisons of AfL and Non-AfL teachers' use of question types will be presented. Second, the types of questions used in each subject will be analysed, drawing comparisons across subjects. Third, the questions used by each of the teachers teaching two different subjects will be compared. Fourth, a comparison between the types of questions used by different teachers in the same subject will be analysed. Next, all the results presented throughout the chapter will be discussed, and finally, the chapter will close with a summary and the main conclusions.

6.1 TEACHERS' QUESTION TYPES ACROSS AFL VS. NON-AFL GROUPS

This section presents the frequency with which each teacher uses each question type and provides a comparison between the types of questions used by AfL teachers versus the types of questions used by Non-AfL teachers. As already explained in Chapter 4 (data and methodology), since teaching units included a different number of sessions, the sessions analysed per teacher varied (16 sessions for Teacher 1, 6 for Teacher 2, 11 for Teacher 3, and 14 for Teacher 4). The following **Table 6.1** shows the raw numbers and percentages of the types of questions asked by the teachers[3], as well as the total number of questions, and the mean number of questions per session. The ratio of number of questions per classroom session is 123.7 for Teacher 1, 62.3 for Teacher 2, 77.2 for Teacher 3, and 48 for Teacher 4. However, the percentages of the teachers' questions regarding the total number of teacher moves reveal that AfL teachers ask more questions (41% in the case of Teacher 1 and 43% in the case of Teacher 2, as opposed to 33% for Teacher 3 and 36% for Teacher 4). The frequency of questioning has been proven to be important, as it promotes student engagement in learning (Crooks 1988; Brophy & Good 1986).

Moving on to question types, *questions for facts/definitions* are the most frequent ones among the four teachers. However, percentages range from 25.7% for Teacher 2 to 86.7% for Teacher 3. This might be explained by the fact that not all teachers enquire about knowledge in the same way: some of them enquire about knowledge as something fixed, stable and objective, while others prefer to ask about content as something under construction, built collaboratively by teacher and students. In the former case, questions for facts will appear more frequently, whereas in the second case, questions for reasons, explanations or opinions might be more numerous. *Questions for opinions* seem to be quite present, with Teacher 2 having the highest percentage (18.2%) and Teacher 3 being the exception, showing a very low frequency (1.6%). *Questions for reasons* seem to be quite rare, with the highest frequency being 6.7% in the case of Teacher 1. In the case of *meta-cognitive questions*, only Teacher 2 uses them on a regular basis (15.8%), but their frequency is very low for the rest of the teachers (they are, in fact, the least frequent). *Questions for explanations/elaboration* are more present in the case of Teachers 1 and 2, who use them twice more often than Teachers 3 and 4. The same can be said for *meta-questions*, which barely appear in the case of Teachers 3 and 4 but have a strong weight in the case of Teachers 1 and 2. As for *language questions*, Teachers 2 and 3 ask them less frequently than Teachers 1 and 4. Within language questions, *content-obligatory questions* are in the majority, which means that the four teachers mainly emphasize language aspects when they are crucial to acquire the content.

[3] The results presented in percentages will be more accurate, since the total number of questions for each teacher varies greatly due to different possible reasons, like the methodology used, the teaching style or the type of subject.

	AfL				Non-AfL			
	Teacher 1		Teacher 2		Teacher 3		Teacher 4	
	%	N	%	N	%	N	%	N
Questions for facts/definitions	41.4%	819	25.7%	96	86.7%	736	60.1%	375
Questions for explanations/elaboration	14.2%	281	17.9%	67	5.7%	48	5.8%	36
Questions for reasons	6.7%	132	2.4%	9	1.3%	11	2.9%	18
Questions for opinions	7.1%	141	18.2%	68	1.6%	14	10.7%	67
Meta-cognitive questions	1.8%	35	15.8%	59	0.4%	3	5.8%	36
Meta-question	16.8%	333	14.4%	54	0.9%	8	0.6%	4
Language question	12.0%	238	5.6%	21	3.4%	29	14.1%	88
content-obligatory	56.3%	134	71.4%	15	86.2%	25	59.1%	52
content-related	43.7%	104	28.6%	6	13.8%	4	40.9%	36
Total number of questions		1979		374		849		634
Mean number of questions per session		123.7		62.3		77.2		48
Mean percentage of questions		41%		43%		33%		36%

Table 6.1 Teachers' question types.

In **Table 6.2**, a comparison between Teachers 1 and 2 (AfL) with Teachers 3 and 4 (Non-AfL) is made in order to compare the two types of methodologies (AfL and Non-AfL), and also because, as we observed in the previous **Table 6.1**, there seem to be more commonalities between Teacher 1 and Teacher 2, on the one hand, and Teachers 3 and 4, on the other. The Chi-square column in the table shows that the differences between the two groups of teachers are significant regarding all the question types. In the case of *meta-cognitive questions* the difference is a bit less significant ($p < 0.05$). The only differences which do not turn out to be significant are those related to *language questions*. The significant differences are the following: AfL teachers ask more *questions for explanations*, *questions for reasons*, *questions for opinions*, *meta-cognitive questions*, *language questions*, and *meta-questions*, whereas Teachers 3 and 4 ask more *questions for facts*.

	AfL		Non-AfL			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	38.9%	915	75.4%	1111	23.57***	485.45***
Questions for explanations/elaboration	14.8%	348	5.7%	84	8.73***	74.68***
Questions for reasons	6.0%	141	2.0%	29	5.90***	34.54***
Questions for opinions	8.9%	209	5.5%	81	3.85***	14.80***
Meta-cognitive questions	4.0%	94	2.6%	39	2.21**	4.90**
Meta-questions	16.4%	387	0.8%	12	15.89***	236.99***
Language questions	11.0%	259	7.9%	117	3.10***	9.60***
content-obligatory	57.5%	149	65.8%	77	1.52	2.31
content-related	42.5%	110	34.2%	40	1.52	2.31
Total		2353		1473		

Note: **$p < 0.05$; ***$p < 0.001$

Table 6.2 Teachers' question types: comparing AfL and Non-AfL teachers.

The results, then, show a common pattern in the types of questions asked by the four teachers: *questions for facts* are the most frequent type of question used by all of them. However, there

are also significant differences in the AfL and Non-AfL teachers' use of question types: Non-AfL teachers ask more *questions for facts* in a significant way (see **Extract 6.1** below), while AfL teachers ask significantly more *questions for explanations* (illustrated in **Extract 6.2**), *for reasons, for opinions* (**Extract 6.3**), *meta-cognitive questions, language questions* (see **Extract 6.4**), and *meta-questions* (illustrated in **Table 4.2**: Classification of episodes and their analysis).

Extract 6.1 Teacher 4 (Non-AfL): question for fact (Citizenship, first unit, class 1)

TCH: Greece is in which sea?
STU: The Mediterranean sea

Extract 6.2 Teacher 2 (AfL): question for explanation (Citizenship, second unit, class 2)

TCH: Do you know how to explain what Guillermo is saying?
STU: Yes, that eh, do a sport is very healthy for your body

Extract 6.3 Teacher 1 (AfL): question for opinion (Citizenship, second unit, class 1)

TCH: What's your opinion? Was it easy or difficult to read?
STU: It was difficult because we read all together it was…
STU: And very long

Extract 6.4 Teacher 1 (AfL): language question (Science, second unit, class 5)

TCH: The tiny muscles is plural, so do you need "do" or "does"?
STU: Do

Extract 6.5 Teacher 2 (AfL): meta-question (Drama, first unit, class1)

TCH: Something negative and something positive, please. Something to improve on and something they did really well
STU: That they did all something at the same time

In sum, despite the fact that *questions for facts* are the most frequent type for both AfL and Non-AfL teachers, the former tend to ask a much wider variety of questions types. In contrast, in the case of Non-AfL teachers, questions other than for facts are very scarce.

6.2 TEACHERS' QUESTION TYPES ACROSS SUBJECTS

This section addresses the types of questions asked by the teachers in each of the four subjects (Science, Citizenship, Arts, and Drama), to see whether the *subject* is a determinant variable when it comes to the types of questions used. The number of sessions analysed are 15 in

Science, 22 in Citizenship, 8 in Arts, and 2 in Drama. Two-way comparisons between *subject* and *question type* have been carried out for all the subjects. Nonetheless, it must be admitted that, as the data does not include the four teachers in the four subjects, it is difficult to discern whether differences are due to *subject* or *teacher*, even with these comparisons. In order to offer further insights in this aspect, analyses of the same teacher teaching two different subjects, and the same subject taught by different teachers, have also been included.

6.2.1 Science lessons

In Science, the most predominant question type is *questions for facts/definitions* (61.6%), followed by *questions for explanations* (12.3%), and *language questions* (9.6%) (see **Table 6.3** below). *Meta-cognitive questions* (0.9%) and *questions for opinions* (3.6%) are the least frequent ones. Regarding *language questions*, their frequency is quite balanced, with *content-obligatory questions* making up 55.8% and *content-related* questions 44.2%. Overall, then, it could be argued that, in CLIL Science classrooms, content tends to be enquired as facts.

	Science	
	%	N
Questions for facts/definitions	61.6%	1108
Questions for explanations/elaboration	12.3%	222
Questions for reasons	4.9%	89
Questions for opinions	3.6%	64
Meta-cognitive questions	0.9%	17
Meta-questions	7.1%	128
Language questions	9.6%	172
content-obligatory	55.8%	96
content-related	44.2%	76

Table 6.3 Question types in Science classrooms.

When Science lessons are compared to Citizenship, significant differences appear. At this point, it is important to take into account that Science is taught by Teacher 1 and Teacher 3, and Citizenship by Teachers 1, 2, and 3. Thus, Teacher 1 is present in the two subjects compared.

	Science		Citizenship			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	61.6%	1108	41.1%	679	12.25***	143.94***
Questions for explanations/elaboration	12.3%	222	11.5%	190	0.75	0.56
Questions for reasons	4.9%	89	4.3%	71	0.90	0.81
Questions for opinions	3.6%	64	12.2%	202	9.68***	91.19***
Meta-cognitive questions	0.9%	17	5.2%	86	7.41***	54.09***
Meta-questions	7.1%	128	14.1%	232	6.70***	44.40***
Language questions	9.6%	172	11.6%	191	1.93*	3.71*
content-obligatory	55.8%	96	64.4%	123	1.67*	2.79*
content-related	44.2%	76	35.6%	68	1.67*	2.79*
Total		1800		1651		

Note: *$p < 0.1$; ***$p < 0.001$

Table 6.4 Comparison question types between Science and Citizenship.

Several question types turn out to be significantly different: a strong difference ($p < 0.05$) appears in *questions for facts*, which are more frequent in Science, while *questions for opinions*, *meta-cognitive questions*, and *meta-questions* are more frequent in Citizenship. A weaker significance ($p < 0.1$) appears in *language questions* and its subtypes. In general, *language questions* are more present in Citizenship, and within *language questions*, *content-obligatory* appear more frequently in Citizenship and *content-related* in Science.

Results obtained from the comparison between Science and Citizenship tell us that, despite some commonalities (such as the similar use of *questions for explanations* and *for reasons*), Citizenship seems to offer more space for exploring and arguing personal opinions (as there are more *questions for opinions* and *meta-cognitive questions*), as well as for reflecting on learning and assessment (as there are more *meta-questions*), whereas Science tends to present knowledge as fact more frequently.

To illustrate these differences, **Extracts 6.6 and 6.7** provide examples of questions in the two different subjects (in fact, taught by the same teacher): in **Extract 6.6**, the teacher is asking a *question for fact* in a Science lesson, whereas in **Extract 6.7**, the teacher is asking a *question for an opinion* in a Citizenship lesson.

Extract 6.6 Teacher 1. Science, question for fact

TCH: So, in order to talk about the leg, what two muscles do you need?
STU: Eh, the quadriceps

Extract 6.7 Teacher 1. Citizenship, question for opinion

TCH: What's your opinion, was it easy or difficult to read?
STU: It was difficult because we read all together it was…

The extracts above come from the same teacher. As will be shown in section 6.3 below, the general differences found between these two subjects (all teachers analysed) are the same as those found in this individual teacher (Teacher 1). Specifically, in Citizenship she uses significantly more *questions for opinions*, *meta-cognitive questions*, and *meta-questions*, and in Science more *questions for facts*.

Finally, as far as language aspects are concerned, although weaker differences are found, there are more *language questions*, and specifically more *content-obligatory questions* in Citizenship.

Turning now to the comparison between Science and Arts, significant differences also appear. As a reminder, Science is taught by Teachers 1 and 3, and Arts by Teachers 3 and 4. Therefore, once again, we have one teacher (Teacher 3) teaching both subjects.

	Science		Arts			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	61.6%	1108	80.4%	217	6.06***	36.08***
Questions for explanations/elaboration	12.3%	222	4.4%	12	3.83***	14.57***
Questions for reasons	4.9%	89	3.7%	10	0.89	0.79
Questions for opinions	3.6%	64	1.5%	4	1.78*	3.18*
Meta-cognitive questions	0.9%	17	1.9%	5	1.36	1.84
Meta-questions	7.1%	128	3.7%	10	2.09***	4.38***
Language questions	9.6%	172	4.4%	12	2.76**	7.57**
content-obligatory	55.8%	96	50.0%	6	0.39	0.15
content-related	44.2%	76	50.0%	6	0.39	0.15
Total		1800		270		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.5 Comparison question types between Science and Arts.

As illustrated in **Table 6.5**, *language questions*, *questions for opinions*, and *meta-questions* are significantly more frequent in Science; in turn, *questions for facts* are significantly more frequent in Arts ($p < 0.01$; except for *meta-questions*: $p < 0.05$). *Questions for facts* are, then, more frequent in Science than in Citizenship lessons, but more numerous in Arts than in Science. The rest of differences between the question types used in Science and Arts are not significant.

To sum up, it could be argued that facts are more predominant in Arts, whereas in Science, even though facts are very present too, students are more involved in explanations and reflections about learning and assessment. As an illustration of the differences just explained, see **Extracts 6.8 and 6.9** below, in which there is an example of a *question for fact* in an Arts lesson (**Extract 6.8**), and an instance of a *question for explanation* in a Science lesson (**Extract 6.9**).

Extract 6.8 Teacher 3. Arts, question for fact

TCH: It's an artistic movie which belongs to which century?

Extract 6.9 Teacher 3. Science, question for explanation

TCH: How did they eat?
STU: They hunt animals

In contrast with the case of Teacher 1, who used different questions in Science and Citizenship, Teacher 3 does not seem to vary the type of questions she asks depending on the subject (see section 6.3 below). Interestingly, Teacher 1 is an AfL teacher and Teacher 3 is a Non-AfL teacher.

Finally, the comparison between question types asked in Science and Drama classes also shows important differences. In this case, none of the teachers taught these two subjects: Teachers 1 and 3 taught Science, while only Teacher 2 taught Drama.

	Science		Drama			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	61.6%	1108	21.0%	22	8.38***	67.78***
Questions for explanations/elaboration	12.3%	222	7.6%	8	1.44	2.08
Questions for reasons	4.9%	89	—	—	0.00	5.45***
Questions for opinions	3.6%	64	19.0%	20	7.63***	56.49***
Meta-cognitive questions	0.9%	17	23.8%	25	16.58***	240.57***
Meta-questions	7.1%	128	27.6%	29	7.53***	55.18***
Language questions	9.6%	172	1.0%	1	2.99***	8.89***
content-obligatory	55.8%	96	100.0%	1	0.88	0.79
content-related	44.2%	76	—	—	0.00	0.79
Total		1800		105		

Note: ***$p < 0.001$

Table 6.6 Comparison question types between Science and Drama.

Significant differences ($p < 0.01$) appear in all the types of questions except for *questions for explanations*. Any comparison with Drama must be made carefully, as the Drama lessons in the corpus are less numerous than the other subjects, as the units were shorter (see for instance that the total number of teacher questions in Science is 1800, whereas in Drama it is only 105). That is why, as pointed out earlier, it is more accurate to look at percentages instead of raw numbers. *Questions for facts*, *for reasons*, and *language questions* are significantly more frequent in the case of Science, whereas *questions for opinions*, *meta-questions*, and *meta-cognitive questions* are significantly more frequent in Drama. This can be explained as in Science students have to learn scientific facts at the same time as they need to understand why these facts occur. On the other hand, students in Drama seem to have a lot of space to express and argue personal

opinions. It also appears that Science gives rise to more opportunities to focus on form (*language questions*) than Drama does.

To illustrate these differences between Science and Drama, **Extract 6.10** below shows a *question for reason* in a Science lesson, while **Extract 6.11** illustrates a *meta-question* in a Drama lesson.

Extract 6.10 Teacher 3. Science, question for reason

TCH: Why are they called the Stone Age and the Metal Ages?
STU: eh… the when- the Stone Age is of.. of <L1 piedra L1> and..

Extract 6.11 Teacher 2. Drama, meta-cognitive question

TCH: Ok, how do flowers make you think of football?

6.2.2 Citizenship lessons

In **Table 6.7** below, the raw numbers and frequencies of question types in Citizenship classrooms are presented. Although *questions for facts* are the most frequent ones (41.1%), there is a lot of room for other questions to appear. The least frequent question types are *questions for reasons* (4.3%) and *meta-cognitive questions* (5.2%). *Questions for explanations* (11.5%), *questions for opinions* (12.2%), *language questions* (11.6%), and *meta-questions* (14.1%) are quite balanced. These results may be indicating that, in Citizenship lessons, apart from content being presented as facts, sometimes students are also encouraged to argue their viewpoints, elaborate ideas, and reflect about their learning process.

	Citizenship	
	%	N
Questions for facts/definitions	41.1%	679
Questions for explanations/elaboration	11.5%	190
Questions for reasons	4.3%	71
Questions for opinions	12.2%	202
Meta-cognitive questions	5.2%	86
Meta-questions	14.1%	232
Language questions	11.6%	191
content-obligatory	64.4%	123
content-related	35.6%	68
Total		1651

Table 6.7 Question types in Citizenship classrooms.

As already pointed out above, Citizenship and Science present commonalities and differences. Among the commonalities, *questions for explanations* and *for reasons* are similarly used. Among the differences, in Science teachers use more *questions for facts* while in Citizenship there are more *questions for opinions*, *meta-cognitive questions*, and *meta-questions*.

The comparison between question types in Citizenship and Arts lessons illustrates many differences, as can be seen in **Table 6.8**. It is important to say that Teacher 4 is the only one teaching both subjects (Citizenship is taught by Teachers 1, 2, and 4; Arts by Teachers 3 and 4).

	Citizenship		Arts			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	41.1%	679	80.4%	217	12.45***	143.60***
Questions for explanations/elaboration	11.5%	190	4.4%	12	3.52***	12.30***
Questions for reasons	4.3%	71	3.7%	10	0.45	0.20
Questions for opinions	12.2%	202	1.5%	4	5.33***	28.03***
Meta-cognitive questions	5.2%	86	1.9%	5	2.41***	5.80***
Meta-questions	14.1%	232	3.7%	10	4.78***	22.57***
Language questions	11.6%	191	4.4%	12	3.54***	12.46***
content-obligatory	64.4%	123	50.0%	6	1.00	1.01
content-related	35.6%	68	50.0%	6	1.00	1.01
Total		1651		270		

Note: ***$p < 0.001$

Table 6.8 Comparison question types Citizenship and Arts.

Differences between the two subjects are strongly significant in all question types except for *questions for reasons*. In the case of Citizenship, the questions which are significantly more frequent are: *questions for explanations*, *for opinions*, *meta-cognitive questions*, *meta-questions*, and *language questions*. In the case of Arts, only *questions for facts* are significantly more frequent. These results seem to indicate that Citizenship, as explained before, is a subject that allows for students to explore and argue their opinions, as well as to reflect on their learning, while in Arts teachers tend to present knowledge as fact. Also, when compared to Arts, teachers in Citizenship ask more *language questions*, therefore there is more focus on form.

As an illustration of a *question for fact* in an Arts lesson, see **Extract 6.8** above. In Citizenship lessons, teachers' types of questions are different, as illustrated in **Extract 6.12** below, in which the teacher is asking a *meta-question*, focusing on what students need to do to get a pass in respect.

Extract 6.12 Teacher 1. Citizenship, meta-question

TCH: <u>What do you need for a <L1 sufi L1> for respect today?</u>

These results are partially replicated when Teacher 4 is compared in these two subjects. There are no significant differences in *questions for explanations*, *meta-cognitive questions*, and the *language questions* subtypes. Significant differences appear in *questions for facts*, *questions for reasons*, *questions for opinions*, and *language questions*. *Questions for facts* and *questions for reasons* are more numerous in her Arts lessons, whereas in Citizenship lessons, *questions for opinions* and *language questions* are more frequent. When Teacher 4 is compared in these two subjects, there are fewer differences in the type of questions used than when subjects are compared irrespective of the teacher.

Differences between Citizenship and Drama are also statistically significant in the majority of types of questions, as illustrated in **Table 6.9** below.

	Citizenship		Drama			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	41.1%	679	21.0%	22	4.11***	16.75***
Questions for explanations/elaboration	11.5%	190	7.6%	8	1.22	1.49
Questions for reasons	4.3%	71	—	—	0.00	4.71**
Questions for opinions	12.2%	202	19.0%	20	2.04**	4.15**
Meta-cognitive questions	5.2%	86	23.8%	25	7.72***	57.68***
Meta-questions	14.1%	232	27.6%	29	3.80***	14.36***
Language questions	11.6%	191	1.0%	1	3.39***	11.43***
content-obligatory	64.4%	123	100.0%	1	0.74	0.55
content-related	35.6%	68	—	—	0.00	0.55
Total		1651		105		

Note: $^{**}p < 0.05$; $^{***}p < 0.001$

Table 6.9 Comparison question types Citizenship and Drama.

Again, there is one teacher (Teacher 2) who taught both Citizenship and Drama. In Citizenship, *questions for facts*, *for reasons*, and *language questions* are more frequent in a significant way. In Drama, this is true for *questions for opinions*, *meta-cognitive questions*, and *meta-questions*. Although, when previously compared to Science and Arts, Citizenship was a subject in which there were more opportunities to talk about opinions and learning, when compared to Drama, we find that there is a heavier presence of facts, and that opinions, arguments and reflection on learning and assessment seem to be more present in Drama classes. As for *language questions*, they are more frequent in Citizenship lessons and, thus, there seems to be a stronger linguistic focus.

Results partially coincide with Teacher 2's types of questions in different subjects, who is the only teacher teaching these two subjects. If **Tables 6.9** and **6.15** are compared, it can be noticed that significant differences also appear when Teacher 2 teaches Citizenship or Drama, specifically in *questions for explanations*, *meta-cognitive questions*, *language questions*, and

meta-questions. The greater presence of *questions for facts* and *for reasons* in Citizenship, and the higher frequency of *questions for opinions* in the case of Drama may be, therefore, due to the subject type.

Extracts 6.13 and **6.14** below illustrate the different types of questions that are found in Citizenship and Drama. In **Extract 6.13**, belonging to a Citizenship class, there is a *language question*. **Extract 6.14**, on the other hand, illustrates a *question for opinion* in a Drama class.

Extract 6.13 Teacher 4. Citizenship, language question

TCH: A noun, similar to free is…? Free…?

Extract 6.14 Teacher 2. Drama, question for opinion

TCH: Did you find that easy or hard?
STU: Easy

6.2.3 Arts lessons

In Arts lessons, the presence of *questions for facts* (80.4%) is very high (in fact, the highest in the four subjects), as shown in **Table 6.10** below. The presence of the rest of the question types is, therefore, very scarce (*questions for explanations* amount to 4.4%; *questions for reasons* total 3.7%; *questions for opinions* make up 1.5%; *meta-cognitive questions* are only 1.9%; and *meta-questions* represent 3.7%). Thus, content in these lessons is mostly enquired as facts. One possible explanation for these results is that the two teachers teaching this subject are Non-AfL teachers and, as it was shown in the first section of this chapter, Non-AfL teachers used more *questions for facts* than AfL teachers. Likewise, this explanation seems highly likely, since *questions for facts* appear more frequently in other subjects taught by Non-AfL teachers when compared to the same subject taught by AfL teachers (see section 6.4 below).

	Arts	
	%	N
Questions for facts/definitions	80.4%	217
Questions for explanations/elaboration	4.4%	12
Questions for reasons	3.7%	10
Questions for opinions	1.5%	4
Meta-cognitive questions	1.9%	5
Meta-questions	3.7%	10
Language questions	4.4%	12
content-obligatory	50.0%	6
content-related	50.0%	6
Total		270

Table 6.10 Question types in Arts classrooms.

When Arts is compared to Drama, very significant differences appear (see **Table 6.11** below).

	Arts		Drama			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	80.4%	217	21.0%	22	12.88***	115.47***
Questions for explanations/elaboration	4.4%	12	7.6%	8	1.23	1.51
Questions for reasons	3.7%	10	—	—	0.00	4.00**
Questions for opinions	1.5%	4	19.0%	20	6.57***	38.94***
Meta-cognitive questions	1.9%	5	23.8%	25	7.53***	49.52***
Meta-questions	3.7%	10	27.6%	29	7.26***	46.40***
Language questions	4.4%	12	1.0%	1	1.66*	2.75*
content-obligatory	50.0%	6	100.0%	1	0.92	0.93
content-related	50.0%	6	—	—	0.00	0.93
Total		270		105		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.11 Comparison question types Arts and Drama.

Questions for opinions, *meta-cognitive questions*, and *meta-questions* appear significantly more often in the Drama lessons. In turn, *questions for facts* are more frequent in Arts. Also, *questions for reasons* and *language questions* are significantly more frequently asked in Arts lessons. Drama, then, appears to be a very good environment for students to develop their opinions and reflections. The differences found between these two subjects can be partly due to the fact that Drama is taught by an AfL teacher and Arts by Non-AfL ones.

To see examples of *meta-cognitive questions* and *questions for opinions* in Drama lessons, see **Extracts 6.11** and **6.14** above. **Extracts 6.15** and **6.16**, respectively, show examples of *questions for reasons* and *language questions*, which are more frequent in the Arts lessons:

Extract 6.15 Teacher 4. Arts, question for reason

TCH: What about America? Why didn't they have this kind of art?

Extract 6.16 Teacher 4. Arts, language question

TCH: What's another word we know for strange?

6.2.4 Drama lessons

Table 6.12 below presents the types of questions asked in Drama lessons (in this case, only used by AfL Teacher 2). As it has been pointed out before, the types of questions asked in these lessons seem to indicate that students are very much engaged with opinions, reasoning, and learning, rather than with facts. In fact, this is the only subject in which *questions for facts* are not the most common type (21%). *Meta-questions* are the most frequent ones (27.6%), followed

by *meta-cognitive questions* (23.8%), *questions for facts* (21%), *questions for opinions* (19%), and *questions for explanations* (7.6%). *Language questions* are barely present (1%), and *questions for reasons* are non-existent.

	Drama	
	%	N
Questions for facts/definitions	21.0%	22
Questions for explanations/elaboration	7.6%	8
Questions for reasons	—	—
Questions for opinions	19.0%	20
Meta-cognitive questions	23.8%	25
Meta-questions	27.6%	29
Language questions	1.0%	1
content-obligatory	100.0%	1
content-related	—	—
Total		105

Table 6.12 Question types in Drama classrooms

To sum up, the results provided in this section show that the type of subject seems to have an effect on what types of questions teachers ask. In **Table 6.13** below, we can see the percentages of each type of question in the four different subjects and the variations of frequency depending on the subject.

	Science	Citizenship	Arts	Drama
Questions for facts/definitions	61.6%	41.1%	80.4%	21.0%
Questions for explanations/elaboration	12.3%	11.5%	4.4%	7.6%
Questions for reasons	4.9%	4.3%	3.7%	—
Questions for opinions	3.6%	12.2%	1.5%	19.0%
Meta-cognitive questions	0.9%	5.2%	1.9%	23.8%
Meta-questions	9.6%	11.6%	4.4%	1.0%
Language questions	7.1%	14.1%	3.7%	27.6%
content-obligatory	55.8%	64.4%	50.0%	100.0%
content-related	44.2%	35.6%	50.0%	—

Table 6.13 Summary of question types in all subjects.

These differences help us portray the different learning environments each academic subject represents. In Science, teachers seem to give importance to scientific facts and their causes and explanations, and there is also room for focusing on form and reflecting on learning, illustrated in the use of *language questions* and *meta-questions*. In Citizenship, teachers seem to be more concerned with students' explanations and opinions, although a focus on language and making students reflect on learning and assessment are also reflected in teachers' questions. Arts lessons seem to be lessons in which the preponderance of facts barely leaves space for other types of questions to appear. Finally, in Drama lessons, students' personal opinions, arguments and thoughts about learning and assessment are encouraged through the use of *questions for*

opinions, *meta-cognitive questions*, and *meta-questions*. However, in order to see whether there are question types characteristic of a subject, regardless of the teacher's style, the next section investigates the use of questions in subjects taught by the same teacher.

6.3 TYPES OF QUESTIONS USED BY THE SAME TEACHER TEACHING TWO DIFFERENT SUBJECTS

This section presents the results obtained regarding teachers' question types when the same teacher is compared teaching two different subjects. In contrast with other studies comparing different subjects, which have had different teachers in each of the subjects (Nassaji & Wells 2000; Black et al. 2004; Hodgen & Marshall 2005), one of the strengths of this investigation is that it includes lessons from different subjects taught by the same teacher. In this way, the variable *teacher* would be completely controlled and it could be discerned whether the *subject* taught can be a significant variable.

Table 6.9 below provides the comparison of the types of questions asked by Teacher 1 in Science and Citizenship. In Science, the most frequent types of questions are *questions for facts* and *for explanations*; in the case of Citizenship, the most frequently asked questions are *questions for facts* and *meta-questions*. It seems that depending on the subject she's teaching, Teacher 1 asks different types of questions: in Science, *questions for facts*, *for explanations*, and *content-related language questions* are significantly more frequent; in Citizenship, *questions for opinions*, *meta-questions*, and *content-obligatory language questions* are significantly more frequent. Only *questions for reasons*, *meta-cognitive* and *language questions* do not present significant differences when compared in the two subjects. These results seem to indicate that facts are more predominant in Science lessons, whereas opinions and student reflections appear to be more important in Citizenship. These results are in line with findings presented in the previous section, when Science and Citizenship lessons were compared irrespective of the teacher: in both cases, *questions for facts* are significantly more frequent in Science, and *questions for opinions* and *meta-questions* are more numerous in the case of Citizenship. As for *language questions*, significant differences appear in the two types (also in line with results in section 6.2, when Science and Citizenship classes were compared): in Citizenship, *language questions* significantly focus more on those aspects of language that are required to learn the content (see **Extract 6.17**), whereas the teacher's focus on language which is not absolutely necessary for the acquisition of the content is more frequent in Science (see **Extract 6.18**).

Extract 6.17 Teacher 4. Citizenship, language question, content-obligatory

TCH: So dictatorship is a noun? Is it a noun?

Extract 6.18 Teacher 4. Science, language question, content-related

TCH: The tiny muscles is plural, so do you need "do" or "does"?

This difference of *language question subtypes* in Citizenship and Science could be explained by the fact that the concepts in Citizenship may be new and more abstract, whereas in Science the concepts are not totally new for the students and the teacher focuses on more geneal linguistic aspects, like the present simple third person singular *–s*.

	Science		Citizenship			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	46.2%	516	35.2%	303	4.94***	24.10***
Questions for explanations/elaboration	16.1%	180	11.7%	101	2.76***	7.62***
Questions for reasons	7.2%	81	5.9%	51	1.17	1.37
Questions for opinions	4.7%	52	10.3%	89	4.90***	23.76***
Meta-cognitive questions	1.3%	15	2.3%	20	1.64	2.70
Meta-questions	11.4%	128	23.8%	205	7.38***	53.10***
Language questions	13.1%	146	10.7%	92	1.61	2.59
content-obligatory	48.6%	71	68.5%	63	3.05***	9.04***
content-related	51.4%	75	31.5%	29	3.05***	9.04***
Total		1118		861		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.14 Comparison question types: Teacher 1 in Science and Citizenship.

In order to illustrate the differences presented in **Table 6.9**, **Extract 6.12** above shows an example of a *meta-question* in a Teacher 1's Citizenship lesson, and **Extract 6.19** instantiates a *question for explanation* in one of the Science lessons taught by Teacher 1.

Extract 6.19 Teacher 1. Science, question for explanation

TCH: How can you explain what the doctor did for María Luisa's sister?

Moving on to Teacher 2, she also uses different question types in her Drama and Citizenship lessons. In Citizenship, the most frequent question types are *questions for facts* and *questions for explanations*; in Drama, the most frequently asked questions are *meta-cognitive questions* and *questions for facts*. The question types that significantly vary depending on the subject (Citizenship or Drama) are: *questions for explanations, for reasons,* and *language questions* significantly more frequent in Citizenship lessons, as opposed to *meta-cognitive questions* and *meta-questions,* with a stronger presence in Drama classes. *Questions for facts, for opinions,*

and the two *subtypes of language questions* do not present any significant differences when Teacher 2 is compared in the two subjects. These results slightly differ with those presented in section 6.2 (see **Table 6.5**): in this case, no significant differences appear in *questions for facts* or in *questions for opinions*, and significant differences are found in *questions for explanations*. However, in general terms, there are similarities when comparing Citizenship and Drama in this particular teacher and the same two subjects including all the teachers in the study. In Citizenship, reasoning and focus on form are more present than in Drama, whereas in Drama the strong presence of *meta-questions* and *meta-cognitive questions* appear to indicate that students' views and reflections are crucial.

	Citizenship		Drama			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	27.5%	74	21.0%	22	1.30	1.70
Questions for explanations/elaboration	21.9%	59	7.6%	8	3.28***	10.52***
Questions for reasons	3.3%	9	—	—	0.00	3.60*
Questions for opinions	17.8%	48	19.0%	20	0.27	0.07
Meta-cognitive questions	12.6%	34	23.8%	25	2.68***	7.09***
Meta-questions	9.3%	25	27.6%	29	4.65***	20.53***
Language questions	7.4%	20	1.0%	1	2.46***	5.99***
content-obligatory	70.0%	14	100.0%	1	0.62	0.42
content-related	30.0%	6	—	—	0.00	0.42
Total		269		105		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.15 Comparison question types: Teacher 2 in Citizenship and Drama.

Extract 6.11 above illustrated *meta-cognitive questions* in Drama lessons, taught by Teacher 2. **Extract 6.20** and **75** below illustrate two question types which are significantly more frequent in Teacher 2's Citizenship lessons than in her Drama lessons: these are *questions for explanations* (**Extract 6.20**) and *language questions* (**Extract 6.21**).

Extract 6.20 Teacher 2. Citizenship, question for explanation

TCH: Who can explain to me what bullying means?

Extract 6.21 Teacher 2. Citizenship, language question

TCH: Helps your body to be relaxed, how do you spell "relaxed"?

As far as Teacher 3 is concerned, when she teaches Science, the most frequently asked questions are *questions for facts* and *for explanations* (similarly to Teacher 1); when she teaches Arts, again *questions for facts* are the most frequent, followed by *meta-questions*. In contrast with Teachers 1 and 2, the types of questions used by Teacher 3 are not so different across subjects (see **Table 6.16**): only *meta-questions* and the *subtypes of language questions* are

significantly different in frequency depending on the subject. In Arts, significantly more *meta-questions* and *content-related language questions* are found (indeed, all *language questions* are *content-related*, and no *meta-questions* are found in Science at all); in Science, significantly more *content-obligatory language questions* are present. Therefore, in Teacher 3's lessons, the only differences found when she's teaching Science or Arts is that there is more space for students' reflections on their learning and assessment in Arts, and that practically all the *language questions* found in Science deal with necessary linguistic aspects, whereas all *language questions* in Arts refer to linguistic features that are not fundamental in order to learn the content. There are a lot of differences, then, between these results and those presented in section 6.2 above. In section 6.2, the differences in the question types used portrayed different types of lessons: Science, in which students' opinions, explanations, reflections on learning and assessment were important, as well as focusing on form; Arts, in which knowledge was presented as fact and students' opinions and arguments had little presence.

	Science		Arts			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	86.8%	592	86.2%	144	0.20	0.04
Questions for explanations/elaboration	6.2%	42	3.6%	6	1.29	1.66
Questions for reasons	1.2%	8	1.8%	3	0.64	0.41
Questions for opinions	1.8%	12	1.2%	2	0.51	0.26
Meta-cognitive questions	0.3%	2	0.6%	1	0.60	0.36
Meta-questions	—	—	4.8%	8	0.00	32.98***
Language questions	3.8%	26	1.8%	3	1.29	1.65
content-obligatory	96.2%	25	—	—	0.00	20.91***
content-related	3.8%	1	100.0%	3	8.36***	20.91***
Total		682		167		
Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$						

Table 6.16 Comparison question types: Teacher 3 in Science and Arts.

Table 6.12 shows that, in line with Teacher 3, Teacher 4's questions are not very different across subjects, when compared to Teachers 1 and 2 (see **Table 6.17**). *Questions for facts* and *language questions* are the most frequent ones in the case of both subjects (Citizenship and Arts). For Teacher 4, the difference between Citizenship and Arts is especially significant when it comes to *questions for facts*, *for reasons*, and *for opinions* (*facts* and *reasons* being frequent in Arts, *questions for opinions* more frequent in Citizenship). The comparison of frequencies of *language questions* and *meta-questions* shows a weaker significance. *Language questions* are more abundant in Citizenship while *meta-questions* appear more frequently in Arts (though the frequency of *meta-questions* is very low in both subjects). When comparing these results to those in section 6.2 (specifically, the comparison of question types in Citizenship and Arts - **Table 6.8**), both similarities and differences are found. Among the similarities, there are more

questions for facts in Arts, and more *questions for opinions* and *language questions* in Citizenship; among the differences, in Teacher 4's lessons, *questions for reasons* are significantly more frequent in Arts whereas there were not significant differences between the subjects when the analysis included all the teachers (see section 6.2, **Table 6.8**). Contrary to general results of subjects, in lessons taught by Teacher 4, *meta-cognitive questions* are not significantly more frequent in Citizenship and *meta-questions* are significantly more frequent in Arts. It could then be argued that, in the case of Teacher 4's Citizenship lessons, students' viewpoints, arguments and reflections have less weight than in Teachers 1 and 2's Citizenship lessons (as it will be shown in the next section).

	Citizenship		Arts			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	58.0%	302	70.9%	73	2.45***	5.98***
Questions for explanations/elaboration	5.8%	30	5.8%	6	0.03	0.00
Questions for reasons	2.1%	11	6.8%	7	2.61***	6.74***
Questions for opinions	12.5%	65	1.9%	2	3.18***	9.96***
Meta-cognitive questions	6.1%	32	3.9%	4	0.90	0.81
Meta-questions	0.4%	2	1.9%	2	1.81*	3.28*
Language questions	15.2%	79	8.7%	9	1.71*	2.93*
content-obligatory	58.2%	46	66.7%	6	0.48	0.24
content-related	41.8%	33	33.3%	3	0.48	0.24
Total		521		103		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.17 Comparison question types: Teacher 4 in Citizenship and Arts

In order to illustrate significant differences, **Extract 6.22** below exemplifies a *question for fact* in a Teacher 4's Arts lesson, and **Extract 6.23** a *question for opinion* in a Teacher 4's Citizenship lesson (see also **Extract 6.13** for an example of a *language question* in a Teacher 4's Citizenship lesson).

Extract 6.22 Teacher 4. Arts, question for fact

TCH: <u>What kind of lines do we get with the ruler?</u>

Extract 6.23 Teacher 4. Citizenship, question for opinion

TCH: <u>What do you think about teachers? In general, what do you think about them?</u>

In this section, results have shown that not all the teachers vary their question types depending on the subject they are teaching. For AfL teachers (Teachers 1 and 2), the type of subject they are teaching seems to affect the types of questions they ask. On the contrary, the school subject does not seem to have such an effect on the question types used by Non-AfL teachers (particularly in the case of Teacher 3). In other words, the frequencies of question types are

quite similar in the case of the Non-AfL teachers (Teachers 3 and 4) teaching two different subjects, as the results reveal few significant differences. Consequently, the differences depicted across subjects in the previous section are especially valid in AfL classrooms.

6.4 TYPES OF QUESTIONS USED BY DIFFERENT TEACHERS IN THE SAME SUBJECT

In this section, different teachers' questions in the same subject are compared. This is another way of comparing teachers' types of questions, with the variable *subject* being controlled. Due to the fact that not all the teachers under analysis taught all the subjects, not all the comparisons are possible. The comparisons that have been feasible are: Teacher 1 and Teacher 3 compared in Science; Teachers 1, 2, and 4 in Citizenship; and Teacher 3 and Teacher 4 contrasted in Arts. We start by comparing Teacher 1 with Teacher 3 in the teaching of Science (**Table 6.13**). There are significant differences in all the types of questions, only *meta-cognitive questions* having a slightly weaker significance. Teacher 1 asks more of all types of questions with two exceptions: *questions for facts*, where Teacher 3 almost doubles Teacher 1's percentage, and *content-obligatory language questions*. This means that, whereas Teacher 3 clearly emphasizes facts, Teacher 1 wants students to look for explanations and reasons. Also, Teacher 1 makes students reflect on their learning and assessment, as opposed to Teacher 3. Significantly more *language questions* are used by Teacher 1, thus there is more negotiation of form than in Teacher 3's Science lessons. Teacher 3 only refers to linguistic aspects when these are important to understand the content; on the contrary, Teacher 1 not only refers to language in those cases, but she also emphasizes form even when it is not crucial for the learning of the content.

	Teacher 1		Teacher 3			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	46.2%	516	86.8%	592	18.80***	295.78***
Questions for explanations/elaboration	16.1%	180	6.2%	42	6.29***	38.72***
Questions for reasons	7.2%	81	1.2%	8	5.82***	33.23***
Questions for opinions	4.7%	52	1.8%	12	3.22***	10.33***
Meta-cognitive questions	1.3%	15	0.3%	2	2.23**	4.98**
Meta-questions	11.4%	128	—	—	0.00	84.06***
Language questions	13.1%	146	3.8%	26	6.55***	41.91***
content-obligatory	48.6%	71	96.2%	25	4.76***	20.21***
content-related	51.4%	75	3.8%	1	4.76***	20.21***
Total		1118		682		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.18 Teachers 1 and 3: comparison types of questions in Science.

Table 6.14 shows the comparison between the question types asked by Teachers 1 and 2 teaching Citizenship. As these two are the teachers in schools implementing AfL, fewer differences are expected. However, significant differences are also found (though not as many as in the previous case). Teacher 1 asks more *questions for facts*, *for reasons*, and *meta-*

questions. Teacher 2, on the other hand, asks significantly more *questions for explanations*, *for opinions*, and *meta-cognitive questions*. These results may indicate that Teacher 2 gives more importance to students' explanations (see **Extract 6.20** above), opinions, and arguments, whereas Teacher 1 focuses more on facts and their reasons, and on making students think about their learning processes (see **Extract 6.12** above).

	Teacher 1		Teacher 2			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	35.2%	303	27.5%	74	2.34**	5.44***
Questions for explanations/elaboration	11.7%	101	21.9%	59	4.22***	17.55***
Questions for reasons	5.9%	51	3.3%	9	1.65	2.71*
Questions for opinions	10.3%	89	17.8%	48	3.31***	10.84***
Meta-cognitive questions	2.3%	20	12.6%	34	7.07***	47.94***
Meta-questions	23.8%	205	9.3%	25	5.22***	26.64***
Language questions	10.7%	92	7.4%	20	1.56	2.43
content-obligatory	68.5%	63	70.0%	14	0.13	0.02
content-related	31.5%	29	30.0%	6	0.13	0.02
Total		861		269		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.19 Teachers 1 and 2: comparison types of questions in Citizenship.

Now, Teacher 1 versus Teacher 4 are compared in Citizenship (see **Table 6.15**). Teacher 1 asks more frequently *questions for explanations*, *for reasons*, and *meta-questions*; in the case of Teacher 4, *questions for facts*, *meta-cognitive questions*, and *language questions* are more numerous. Only *questions for opinions* and the *subtypes of language questions* show no statistically significant differences. In Citizenship lessons taught by Teacher 4, then, knowledge as fact, students' argumentation, and negotiation of form have a stronger presence than in the Citizenship lessons taught by Teacher 1, where there are significantly more explanations, reasons, and reflections on learning and assessment.

	Teacher 1		Teacher 4			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	35.2%	303	58.0%	302	8.48***	68.40***
Questions for explanations/elaboration	11.7%	101	5.8%	30	3.69***	13.49***
Questions for reasons	5.9%	51	2.1%	11	3.33***	11.01***
Questions for opinions	10.3%	89	12.5%	65	1.22	1.50
Meta-cognitive questions	2.3%	20	6.1%	32	3.63***	13.07***
Meta-questions	23.8%	205	0.4%	2	12.47***	139.87***
Language questions	10.7%	92	15.2%	79	2.45***	6.00***
content-obligatory	68.5%	63	58.2%	46	1.39	1.93
content-related	31.5%	29	41.8%	33	1.39	1.93
Total		861		521		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.20 Teachers 1 and 4: comparison types of questions in Citizenship.

To illustrate the significant differences between Teacher 1 and Teacher 4 when asking questions in Citizenship, there is an example of a *question for explanation* in Teacher 1's Citizenship

lesson in **Extract 6.24**, and an example of a *question for fact* in Teacher 4's Citizenship lesson in **Extract 6.25**.

Extract 6.24 Teacher 1. Citizenship, question for explanation

TCH: How did we solve the problem?

Extract 6.25 Teacher 4. Citizenship, question for fact

TCH: So in our democracy, who is able to vote in Spain?

When comparing Teacher 2 and 4 in Citizenship, significant differences are expected to appear as well, as their profiles are different (Teacher 2 is an AfL teacher, and Teacher 4 is a Non-AfL one) and, as shown in **Table 6.2** in section 6.1, there were indeed significant differences between the AfL and the Non-AfL teachers. *Questions for facts* and *language questions* are the only two types which are found to be more numerous for Teacher 4 in a significant way. Teacher 2 asks significantly more *questions for explanations*, *for opinions*, *meta-cognitive questions*, and *meta-questions*. The only question types which do not show statistically significant differences are almost the same as those found in the comparison between Teacher 1 and Teacher 4, that is, *questions for reasons* and the *subtypes of language questions*. The differences found suggest that Citizenship lessons taught by Teacher 2 are not so much about facts but about explanations, opinions, argumentation, and trying to make students aware of learning and assessment processes. In contrast, Teacher 4's Citizenship lessons places more emphasis on facts and on linguistic matters.

	Teacher 2		Teacher 4			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	27.5%	74	58.0%	302	8.47***	65.97***
Questions for explanations/elaboration	21.9%	59	5.8%	30	7.01***	46.43***
Questions for reasons	3.3%	9	2.1%	11	1.05	1.10
Questions for opinions	17.8%	48	12.5%	65	2.04**	4.17**
Meta-cognitive questions	12.6%	34	6.1%	32	3.14***	9.78***
Meta-questions	9.3%	25	0.4%	2	6.71***	42.66***
Language questions	7.4%	20	15.2%	79	3.12***	9.67***
content-obligatory	70.0%	14	58.2%	46	0.96	0.93
content-related	30.0%	6	41.8%	33	0.96	0.93
Total		269		521		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.21 Teachers 2 and 4: comparison types of questions in Citizenship.

See **Extracts 6.23 and 6.25** above for examples of Teacher 4's *language question* and *question for fact*, respectively, in Citizenship lessons. **Extract 6.26** is an illustration of a *question for opinion* asked by Teacher 2 in a Citizenship lesson.

Extract 6.26 Teacher 2. Citizenship, question for opinion

TCH: What did you think of the poem?

Since significant differences were found when comparing separately Teacher 1 and 2 with Teacher 4 in Citizenship lessons, significant differences are also expected to appear in the comparison between Teachers 1 and 2, on the one hand, and Teacher 4 on the other. **Table 6.17** presents these results and confirms the expectations. Teacher 4 asks significantly more *questions for facts*, *language questions*, and *meta-cognitive questions* than Teachers 1 and 2 together. Teachers 1 and 2, however, ask more *questions for explanations*, *for reasons*, and *meta-questions*. *Questions for opinions* and the *subtypes of language questions* present no differences. Again, we find that Citizenship lessons taught by Teacher 4 have a strong orientation to facts and linguistic aspects, whereas in Citizenship lessons taught by Teachers 1 and 2, explanations, reasons, and monitoring of learning and assessment are more important.

	Teacher 1 & 2		Teacher 4			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	35.0%	365	46.4%	204	4.12***	16.80***
Questions for explanations/elaboration	18.5%	193	14.5%	64	1.85*	3.41*
Questions for reasons	6.5%	68	2.5%	11	3.16***	9.94***
Questions for opinions	16.8%	175	16.4%	72	0.20	0.04
Meta-cognitive questions	3.6%	37	6.1%	27	2.24**	5.00**
Meta-questions	10.5%	109	—	—	0.00	49.68***
Language questions	9.1%	95	14.1%	62	2.85***	8.08***
content-obligatory	66.3%	63	54.8%	34	1.45	2.09
content-related	33.7%	32	45.2%	28	1.45	2.09
Total		1042		440		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.22 Teachers 1/2 and 4: comparison types of questions in Citizenship.

The last comparison to be made is between Teacher 3 and Teacher 4 teaching Arts (**Table 6.18**). Both teachers belong to the same type of school (not-AfL), so no major differences are expected. At the highest level of significance only appear *questions for facts* and *language questions*, the former being more frequent in the case of Teacher 3 (see **Extract 6.8** as an illustration), the latter in the case of Teacher 4. At the medium level of significance ($p < 0.05$) we have *questions for reasons* (more numerous in Teacher 4's Arts lessons, see **Extract 6.15** for an example) and *language questions subtypes* (*content-obligatory* are more recurrent in Teacher 4's lessons, whereas *content-related* are more common in Teacher 3's). Finally, at the lowest level of significance ($p < 0.1$), there are *meta-cognitive* questions, which, although scarce, are asked more frequently by Teacher 4. Results, then, appear to indicate that Arts lessons taught by Teacher 3 emphasize the learning of facts, while Arts lessons taught by Teacher 4 do not only focus on facts but also on explanations, reasons, and argumentation.

Teacher 4 also focuses more on language, including both language necessary and not necessary for the learning of the content, as opposed to Teacher 3, who only asks for linguistic aspects which are not fundamental for content learning and understanding.

	Teacher 3		Teacher 4			
	%	N	%	N	*T*	χ^2
Questions for facts/definitions	86.2%	144	70.9%	73	3.13***	9.52***
Questions for explanations/elaboration	3.6%	6	5.8%	6	0.86	0.75
Questions for reasons	1.8%	3	6.8%	7	2.12**	4.47**
Questions for opinions	1.2%	2	1.9%	2	0.49	0.24
Meta-cognitive questions	0.6%	1	3.9%	4	1.95*	3.78*
Meta-questions	4.8%	8	1.9%	2	1.20	1.45
Language questions	1.8%	3	8.7%	9	2.72***	7.23***
content-obligatory	—	—	66.7%	6	0.00	4.00`**
content-related	100.0%	3	33.3%	3	2.24**	4.00**
Total		167		103		

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

Table 6.23 Teachers 3 and 4: comparison types of questions in Arts.

The results presented in this section show that significant differences in the use of question types also appear when we compare different teachers teaching the same subject (significant differences have appeared in all the comparisons). Therefore, one might interpret that not only the *subject* is an important variable to take into account when teachers' question types are analysed (as shown in section 6.2 and 6.3) but the variable *teacher* is also affecting the use of questions. The difference between an AfL and Non-AfL teacher seems to be the determining one (as proven in this chapter), as AfL teachers tend to vary their question types according to the subject they are teaching, as opposed to Non-AfL teachers, who seem to remain more constant in their use of questions in different subjects. However, other teaching style factors (such as experience, for instance) might also be affecting this use of question types.

6.5 DISCUSSION

Questioning is one of the most important resources in AfL (Black et al. 2003; Wiliam et al. 2004; Black et al. 2004; Torrance & Pryor 1998; Black & Wiliam 1998a, b; Harlen & Winter 2004; Harrison & Howard 2009). AfL researchers have often claimed that teachers' questions should encourage student participation and thinking, and elicit more than one-word correct answers (Harlen & Winter 2004; Black & Wiliam 1998b). A question is not formative on its own, a question is formative when it reveals student understandings and promotes discussion and collaboration. This necessarily involves long and complex students' responses, both in

terms of subject content and language use (Harrison & Howard 2009). Questions are also important because they are one of the most frequent diagnostic skills used by teachers in primary foreign language classrooms when assessing students' language ability (Edelenbos & Kubanek-German 2004). In the context of CLIL, this becomes very relevant, as the duality content—language is always present, and therefore, diagnosis needs to be done at two different levels: that of language and that of content.

The frequency of questions has been correlated with positive results in student achievement, probably because they engage students in learning, consolidating it, and offer practice on the material (Crooks 1988; Brophy & Good 1986). In the present study, the frequency of questions per total number of moves has shown that AfL teachers are the ones who more frequently ask questions, although differences are not significant. The ratio of number of questions per teacher is 123.7 for Teacher 1, 62.3 for Teacher 2, 77.2 for Teacher 3 and 48 for Teacher 4. One possible explanation for Teacher 1 to display so many questions is that the subjects she teaches (Science and Citizenship) are the ones in which more questions are asked, as opposed to Arts and Drama (see **Tables 6.3**, **6.7**, **6.10**, **6.12**). The fact that Teachers 2, 3, and 4 asked fewer questions could be explained by the fact that all three taught either Drama or Arts, which displayed a lower frequency of questions (see **Tables 6.3**, **6.7**, **6.10**, **6.12**). It could be argued that Citizenship and Science are the two subjects with the greatest number of initiations because Teacher 1, who teaches both of them, is the one who asks the highest number of questions. However, if we take a closer look at the number of questions asked by Teachers 2, 3, and 4 in Science (Teacher 3) or Citizenship (Teachers 2 and 4) and compare them to the number of questions they ask in Arts (Teachers 3 and 4) or Drama (Teacher 2), it is clear that the frequency of questions in either Science or Citizenship outnumber greatly those found in Arts or Drama. Therefore, it seems that the type of subject has an effect on the frequency of questions asked by teachers, regardless of individual teacher styles.

Teachers' use of different types of questions could have important effects for students to be able to work on content and language from different perspectives. Indeed, a variety of questioning techniques is advocated by some researchers (Carlsen 1991; Lyster 2007; Black & Wiliam 1998a; Black & Wiliam 2009; Harrison & Howard 2009; Alexander 2004), although in real classrooms this is difficult to achieve (Black & Wiliam 1998a). If questioning techniques are varied, students are maintained active, which is crucial for learning and AfL (Wragg & Brown 2001). All the lessons observed in this study portray a variety of question types but, of course, the frequency varies across teachers. Teacher 3, for instance, shows the lowest variety in the type of questions she asked, because 86.7% of her questions are *questions for facts/definitions*. The high frequency found for *questions for facts* (especially in Teacher 3's and 4's lessons) has

also been reported in many previous studies. Dalton-Puffer (2007) found that in CLIL Austrian classrooms 89% of the questions were *questions for facts*, a very similar percentage to that found for Teacher 3 (*questions for facts* are also the most frequent type for Teacher 4 (60.1%), though the percentage is not so high). Other studies have shown similar findings. Wragg and Brown (2001) found that 70-80% of teachers' questions in L1 contexts focused on low order skills (remembering facts); Airasian (1997) claimed that 90% of teachers' questions in L1 classes were only aimed at discerning correct responses from incorrect ones; Nystrand & Gamoran (1997), again in L1 contexts, also found a majority of what they call "test questions", which required pre-specified answers which allow teachers to be in control of the interaction and its direction, to the detriment of authentic and more dialogic types of questions; and Ruiz-Primo and Furtak (2006) found questions for definitions to be the most frequent in L1 Science classrooms, many of which were yes/no questions, whereas questions promoting argumentation were quite infrequent. One of the reasons that might explain this heavy presence of *questions for facts* is that teachers do not like waiting for students' responses for too long (Rowe 1974), and the only type of question that can receive immediate answers is *questions for facts* (Harlen & Winter 2004; Black & Wiliam 1998b). As it will be explained in the next chapter, an excessive use of *questions for facts* may limit students' participation and engagement in classroom discourse.

Moving back to the present study, all question types show significant differences when Teachers 1 and 2 are compared to Teachers 3 and 4 (see **Table 6.2**). That is, when AfL teachers are compared to Non-AfL ones. As a consequence, one could argue that the type of programme (AfL or Non-AfL) affects the use of different question types. These results support hypotheses 1 and 2, namely, that there are significant differences in the types of questions asked by teachers in AfL schools compared to those in Non-AfL schools. In fact, the only question type that Teachers 3 and 4 ask more frequently is questions for facts. For the rest, Teachers 1 and 2 have higher percentages. *Meta-cognitive questions* and *questions for reasons*, which are especially important if deep learning is to be triggered and rote memorization is to be avoided (Harlen & Winter 2004), are significantly more frequent in AfL classes. *Meta-questions* were expected to appear only in AfL classes, as it is a very important element of Assessment for Learning. Although *meta-questions* have also been found in Non-AfL classes, the difference between the two types of programmes is statistically significant. Specifically, *meta-questions* just amount to 0.8% of the questions asked in the Non-AfL classes, as opposed to 16.4% in the AfL schools. These results show that in AfL classes students are given more chances of understanding their learning process (Pintrich 2002). As *meta-questions* are crucial if learners are to be initiated into self-assessment, they are a starting point for students to begin reflecting on their learning and

how they can improve it (Wragg & Brown 2001). Again, these findings, then, support hypotheses 1 and 2, which state that AfL discourse is constructed through the use of specific types of questions, and that there will be significant differences in the types of questions (and feedback) used by AfL teachers when compared to Non-AfL ones.

The implementation of AfL not only seems to promote *meta-questions* but also *questions for opinions*, *for reasons*, and *meta-cognitive questions* (which foster students' thinking, reasoning, and understanding of mental processes). Different types of questions have different cognitive demands for students (Bloom et al. 1956; Hardman et al. 2003; Dalton-Puffer 2007). *Questions for reasons*, *for opinions*, and *meta-cognitive questions* are very good at eliciting students' understandings and challenging their thinking (Ruiz-Primo & Furtak 2006; Pintrich 2002). In the case of Non-AfL schools, it seems to be the case that questions which entail low cognitive demand are more frequent (results which coincide with Dalton-Puffer 2007; Stiggins et al. 1989; Slavin 1991; Wragg & Brown 2001; Alexander 2004). Crooks, after a review on the relationship between the cognitive demand of the question and student learning (1988), concludes that high order questions foster achievement (see also Redfield & Rousseau 1981) (and also interest, learning, retention, development of thinking skills), although training and practice with them is needed. Nystrand & Gamoran (1997), however, warn that *factual questions* are not always ineffective, as well as discussion and authentic questions do not always lead to learning. Alexander (2004), who advocates for dialogic teaching, argues that questions need to promote thoughtful and reflective answers. She associates different types of questions with different modes of teaching, and contrasts recitation (in which *questions for facts* are frequent and therefore the cognitive challenge is low) with discussion or scaffolded dialogue (in which different types of thinking and arguing questions appear to construct a more dialogic kind of teaching).

As far as *language-related questions* are concerned, these acquire particular relevance in CLIL contexts. Dealing with language aspects spontaneously in content-based instruction, be it through language questions or through feedback focusing on language (see Chapter 8), is said to be more effective than working on language aspects in language Arts lessons (Lyster 2007). The amount of *language questions* or teacher strategies focusing on language (see also Chapter 8) gives an idea of negotiation of form, which cannot always be separated from negotiation of meaning (Gass 1997), though both are fundamental for language learning (Lyster 2007; Swain 1995). As shown in **Table 6.2** above, AfL teachers also ask more *language questions* than Non-AfL ones, which means there is more negotiation of language in AfL CLIL classrooms. Regarding the types of language questions asked, differences between teachers in AfL schools and Non-AfL schools are not significant. However, the tendency is a more balanced distribution

between *content-obligatory* and *content-related language questions* in AfL lessons, while in Non-AfL schools, on the other hand, it seems that the tendency is for teachers to ask more *content-obligatory language questions*. This may be related to the fact that the type of teaching in Non-AfL schools may be less dialogic and spontaneous, and teachers may be more reticent to depart from what was planned.

Lyster argues (2007) that, in immersion contexts, negotiation of form almost only happens when comprehensibility is needed. And even in those cases, teachers can become experts at interpreting children's interlanguage and accept a minimum level to satisfy communicative needs, which may detriment their L2 development (Lyster 2007). In the case of the present study, teachers started negotiations of form not only when comprehensibility was at stake, but also when grammatical errors were made or when the teacher wanted to emphasize certain linguistic features (see **Extract 6.27**).

Extract 6.27 Teacher 1 (Science, second unit, class 3)

TCH: Yasmina wrote "skeleton is". Should we write "the skeleton is" or the "skeleton are"?
STU: "Is", because we have one

This different approach to the L2, compared to the observations in French immersion contexts pointed out above, might be related to the fact that in the CLIL contexts analysed in this study, primary school teachers are both EFL and content teachers, and thus, they are more aware of language issues when they teach content than content teachers in French immersion classrooms. In the present study, this language awareness was especially noticed in AfL lessons. In addition, the combination of questions that were mainly found in AfL classes, those which look for students' thinking, reasoning, and argumentation together with language questions, is not only crucial for AfL methodology but also to satisfy the needs of CLIL: content and language integrated learning, as both (higher-order) content and language objectives are covered.

In spite of the clear differences between AfL and Non-AfL teachers, this study has shown that the types of questions asked can also be related to individual teaching styles. In the comparison of the same subject taught by different teachers (Teachers 1 and 3 in Science, Teachers 1, 2 and 4 in Citizenship and Teachers 3 and 4 in Arts), significant differences appeared, even when two teachers from the same group (AfL school or Non-AfL school) were compared (Teachers 3 and 4 in Arts; Teachers 1 and 2 in Citizenship). This suggests that, apart from their AfL/Non-AfL methodological approach, these teachers' individual teaching styles also played a role.

Regarding the distribution of questions in different subjects, all question types are present in every subject, but the distribution is different. The most frequent type of question in all subjects, except in Drama, is *questions for facts/definitions*. This type is especially frequent in Arts (80.4%) but also in Science (61.6%). The results on the most frequent types of questions found in Science vary across studies. In line with this study, Ruiz-Primo and Furtak (2006) found that questions asking for definitions were very frequent, as opposed to questions promoting argumentation. In contrast, Nassaji and Wells (2000) found that factual questions in the L1 Science classes that they analysed were less frequent than other question types. To explain this contrast, it may be necessary to take into account diverse classroom cultures and the fact that different geographical contexts have different idiosyncrasies when it comes to education.

Turning back to the present study, *questions for explanations* were frequent in Science in the first place (12. 3%), and Citizenship (11.5%) in the second. This result is expected as explaining scientific events is one of the main objectives in Science (Dale & Tanner 2012; Duschl & Gitomer 1997). The percentage decreases in Drama, and especially in Arts, where only 4.4% of the questions are for explanations. Again, *questions for reasons* are more numerous in Science and Citizenship, and they do not even appear in Drama. It is very common that students have to deal with reasons in Science, both in CLIL and in L1 contexts (Dale & Tanner 2012; Duschl & Gitomer 1997). *Questions for opinions* are one of the most frequent types in Drama, and also very recurrent in Citizenship, possibly because personal matters are more present in these subjects (Dale & Tanner 2012). On the other hand, they are scarce in Science and in Arts (especially in the latter). *Meta-cognitive questions* are another type found to be very frequent in Drama, possibly due to the importance of creativity and personal emotions (Dale & Tanner 2012), scarce in Citizenship and almost non-existent in Science and Arts. *Meta-cognitive questions* might be more frequent in Drama and Citizenship because *questions for opinions* are more frequent in these subjects too, and they are intimately related: it is very common that, after a question for opinion, the teacher asks a metacognitive question, such as “why do you think that?”. The weak presence of *meta-cognitive questions* in Science and Arts could lead us to think that these subjects are more about facts and less about personal opinions. *Language questions* appear more frequently in Citizenship and Science, but they are infrequent in Arts and Drama. This could be explained by the type of subject, as Citizenship, and especially Science deal with more complex concepts, and therefore a focus on the necessary language to express them becomes paramount. Finally, and surprisingly enough, *meta-questions* are mostly used in Drama classes followed by Citizenship, Science, and Arts in the last place. The absence of *meta-questions* in Arts can be influenced by the fact that only Non-AfL teachers teach Arts in the corpus. Likewise, it is possible that the high presence of *meta-questions* in Drama is

influenced by the teacher: Teacher 2, who is the only one that teaches Drama in this corpus, is also the one who uses a higher number of *meta-questions*. Finally, regarding language questions, *content-obligatory* and *content-related*, there are only significant differences when comparing Citizenship and Science, with *content-obligatory* being more frequent in the former. As explained elsewhere, this might be explained by the fact that concepts in Citizenship are new and abstract. In Science, however, language questions were more frequently *content-related* because concepts have been presented and language questions hence deal with more formal aspects, such as subject-verb concordance.

All in all, the use of the different question types in each subject depicts different learning contexts. In Science, facts and their possible explanations/reasons are very relevant, along with negotiation of form and reflection on learning processes and assessment. In Citizenship, students' opinions and explanations are more emphasized, and negotiation of form and students' thinking about learning are also present. Arts lessons seem to give space only to facts, with little space for reasons, arguments and reflections. Finally, in Drama, students' opinions, viewpoints, and arguments are actively encouraged.

However, as pointed out above, all these significant differences may be influenced by significant differences found between teachers. In order to make sure that the differences observed across subjects were not due to differences in individual teacher styles, the questions used by the same teacher in different subjects were compared. In the case of Teacher 1, who is an AfL teacher, significant differences appear when the types of questions she asks in Science are compared to the ones she asks in Citizenship. Therefore, the type of subject seems to be an important factor for her types of questions: *questions for facts, for explanations,* and *content-related language questions* are more abundant in Science, whereas *questions for reasons, meta-questions,* and *content-obligatory language questions* are significantly more frequent in Citizenship. For Teacher 2, the subject also seems important: Citizenship triggers more *questions for explanations, for reasons,* and *language questions*, whereas Drama gives more room to *meta-questions* and *meta-cognitive questions*. This may be explained because in Drama emotions are dealt with and also evaluating peers' and one's work is fostered (Dale & Tanner 2012). Both Teacher 1's and Teacher 2's results greatly coincide with the general findings across subjects. In contrast, for Teacher 3, significant differences only appear in *meta-questions* and *language questions subtypes*: Arts triggers more *meta-questions* and *content-related language questions* (the former maybe due to the fact that in Arts it is important to evaluate artworks). Since Teacher 3's types of questions do not change much from one subject to the other, these results do not coincide with the general results obtained across subjects. Finally, in the case of Teacher 4, the subject also determines, to a certain extent, the frequency of question

types: *questions for opinions*, *meta-cognitive questions*, and *language questions* are significantly more frequent in Citizenship, while in Arts *questions for facts*, *for reasons*, and *meta-questions* are statistically more numerous than in Citizenship. Again, Teacher 4's use of questions is similar to the general distribution of question types across subjects. This result proves another hypothesis of the present study (hypothesis number 3): that differences will be found in the discourse of the same teacher teaching different subjects. This hypothesis was based on the evidence that AfL has been found to vary depending on different factors, one of them being that the same teacher can invoke different pedagogic assumptions and principles in different subjects and with different students (Leung 2004; Torrance & Pryor 1998; Black et al. 2004; Wiliam 2006; Black & Wiliam 1998a; Black & Wiliam 2009). In addition, it is expected that, since different subjects require different genres, different subjects will demand different question types (Llinares & Whittaker 2009; Whittaker & Llinares 2009; Llinares et al. 2012; Dalton-Puffer 2011; Alexander 2004; Wiliam 2006). The link between AfL and the different needs that emerge from different subjects and genres seems to work, as Teachers 1 and 2 appear to be more aware of different subject requirements because they use different questions across subjects. However, it is not so clear in the case of Non-AfL teachers, since their questions are quite homogeneous (especially in the case of Teacher 3) regardless of the type of subject they are teaching.

6.6 SUMMARY AND CONCLUSION

This chapter has focused on the question types asked by CLIL teachers in primary classrooms across subjects. Significant differences appear in every comparison carried out. First, AfL teachers were compared to Non-AfL teachers, and significant differences were found, with Non-AfL teachers asking more *questions for facts* and AfL teachers asking more *questions for explanation*, *for opinions*, *for reasons*, *meta-cognitive questions*, *language questions*, and *meta-questions*. As a consequence, discourse in AfL classrooms seems to be more dialogic and oriented to learning. Next, significant differences also came out of the comparison across subjects, portraying different learning environments for the different subjects (some subjects – Arts and Science – focusing more on facts, others – Citizenship, Drama – more on students' thinking and ideas). Then, the use of question types by the same teacher teaching two different subjects was compared. In this case, significant differences also appeared, but mainly in the case of AfL teachers, who seem to be more sensitive to the different genres required by different subjects. Finally, the use of types of questions by different teachers in the same subject was analysed, producing significant differences not only when an AfL teacher was compared to a

Non-AfL one, but also when teachers from the same group were compared. This last comparison could point to individual teaching style as a possible factor which may also have an influence on the use of question types.

Two main conclusions can be drawn from these results. In the first place, AfL teachers, no matter the subject, seem to ask a wider variety of question types, which make classroom interaction freer and more dialogic, as they do not seem to be always looking for a correct response but let students explore their own ideas and learning (Barnes 1975; Alexander 2004; Black et al. 2003). This is done through questions such as *meta-questions, meta-cognitive questions*, and *questions for opinions*. Secondly, as hypothesized, both type of subject and teacher style are decisive factors that have an effect on the type of questions asked. Subsequent chapters on students' responses and teachers' feedback will provide further information as to what type of classroom discourse takes place in AfL and Non-AfL classes.

7

Results III — Student responses and initiations

This chapter focuses on students' participation in classroom interaction. Two types of participation will be measured: *students' responses* to teachers' questions, and *students' initiations*. Students' responses will be analysed in relation to the questions asked by the teachers (see Chapter 6). Students' initiations will inform about the chances students have to initiate discourse and pursue their own ideas, hence generating an exchange where the teacher is not the initiator but the respondent.

In the first section, students' responses in AfL and Non-AfL classes will be compared in order to discover any significant differences between them. Likewise, there will be comparisons within types of programmes, that is, students' responses from AfL group 1 will be compared to those from AfL group 2, and students' responses from Non-AfL group 3 will be compared to those from Non-AfL group 4. This will be done to prove the homogeneity of each type of programme, or in other words, to make sure that significant differences between groups of the same type of programme do not appear. The second section of the chapter will analyse the type of response triggered by each of the question types in order to see if the complexity of the response is related to the type of question. Thirdly, the types of responses given by students will be examined across subjects, to see if this variable has an effect on the complexity of students' responses. The fourth section will be devoted to students' initiating turns, comparing AfL with Non-AfL groups in relation to the number and the type of initiations. All these sections will combine quantitative with qualitative accounts of the data. Next, some discussion of the results will follow, and finally, the chapter will close with a summary of the main results and the conclusions that can be drawn from them.

7.1 STUDENTS' RESPONSES IN AFL AND NON-AFL CLASSES

In this section, I am going to compare the types of responses to teachers' questions given by students from the different groups. In order to measure the complexity of students' responses, they were classified into: *minimal responses* (yes/no answer), *T-units* (main clause plus any subordinate clauses), and *truncated responses* (incomplete T-unit). Within T-unit responses, there is a further sub-classification: *one-phrase*, *one-clause*, and *more-than-one clause* T-units. **Table 7.1** below compares students' responses from the AfL groups with those from the Non-AfL groups.

	AfL		Non-AfL			
	%	N	%	N	*T*	χ^2
Minimal response	16.0%	544	15.4%	336	0.58	0.34
T-Unit response	77.8%	2641	79.5%	1729	1.49	2.22
Truncated response	4.5%	153	3.9%	85	1.08	1.17
Total		3394		2175		
T-Unit						
one-phrase	38.2%	1296	47.9%	1042	7.20***	51.45***
one-clause	27.1%	920	23.2%	504	3.29***	10.78***
more than one-clause	11.8%	399	7.8%	170	4.74***	22.43***
Total		3394		2175		
Note: ***$p < 0.001$						

Table 7.1 Comparison of students' responses in AfL vs. Non-AfL schools.

It turns out that no significant differences come up in the first classification (minimal response/T-unit/truncated response) across both types of schools. The majority of students' responses are *T-units*, followed by *minimal responses*, and very few *truncated responses*. However, differences appear when we look at T-unit responses more closely. Students' answers in AfL schools are more complex, as there are more *one-clause* (27.1% as opposed to 23.2%) and *more than one-clause responses* (11.8% compared to 7.8%); on the other hand, almost half of T-unit responses in Non-AfL schools are *one-phrase* (47.9%). As will be shown in more detail in further sections of the chapter, we can relate these findings to teachers' question types. The results shown in the previous chapter revealed that teachers in Non-AfL schools asked significantly more questions for facts (which require less complex answers), whereas teachers in AfL schools asked significantly more questions for explanations, for reasons, and meta-cognitive questions (which seem to trigger more complex responses on the students' part).

Extracts 7.1 and 7.2 below will help illustrate these differences. **Extract 7.1** corresponds to a Science lesson from one of the Non-AfL groups in which students are learning about fish.

Extract 7.1 Non-AfL-3, Science, first unit, fourth class

1	TCH:	[Where].. wait a second. Where do they live? Where do we find fish? Nata- eh.. Yasmina
2	STU1:	in the water
3	TCH:	In.. in which wa- in which water?
4	STU1:	In.. in
5	TCH:	Some fish live, [Noelia]?
6	STU1:	[<L1 agua dulce L1>]
7	STU2:	in.. in the sea
8	TCH:	In the ocean
9	STU2:	and others in the river
10	TCH:	And others in rivers, very good.
11	TCH:	Now, who can tell me names of fish that live in the ocean? Andrea ((she says no with her head)) No?? Think about it. Names of fish that live in the ocean
12	STU3:	sardine?
13	TCH:	Sardines. Sardines ((writing it on the board)), good. David
14	STU4:	<L1 monjita L1>
15	TCH:	Mm?
16	STU4:	<L1 monjita. Mon-ji-ta.L1> I don't know what is eh.. s- in English
17	TCH:	I don't know that fish in Spanish either. <L1 ¿Monjita? L1>
18	STU4:	<L1 Es así mu.. muy… L1>
19	TCH:	Think of another fish
20	STU4:	<L1 muy chico, yo qué sé.. Lubina L1>
21	TCH:	Bass ((writing it)). <L1 Lubina L1>. Do we know any other fish..?
22	STU:	Yes
23	TCH:	Think about it.. that lives in the ocean?
24	STU:	Yes
25	TCH:	Eh.. eh.. Gabriela
26	STU5:	eh <L1 Pez payaso L1>
27	TCH:	Goldfish, and you are for-.. I don't know.. how can you forget..? ((writing it on the board))
28	STU:	<L1 Sí, es que no sé cómo se dice L1>
29	TCH:	Goldfish, yes. No, goldfish... It's clownfish.
30	STU:	<L1 Sí L1>, clownfish
31	TCH:	Clownfish is a different one. Clownfish ((writing it)) and a really big, big fish that lives in the ocean, Omar?
32	STU6:	eh.. yeah, squid
33	TCH:	Squid?
34	STU:	<L1 Invertebrado L1>
35	TCH:	Squid.. squids.. Are squids fish?
36	STU:	((Some)): no
37	STU:	Are invertebrates
38	TCH:	What are they?
39	STU:	((Some)): Invertebrates

This is a typical exchange in the classes taught by Teacher 3. In this extract, all her questions are questions for facts (see, for instance, turns 1, 3, 11). In fact, as illustrated in Chapter 6, 86.7% of this teacher's questions were questions for facts/definitions. In the response turn, most of the students' answers are *one-phrase T-unit* responses ("*in the water*", "*sardine*",

"*invertebrates*"…). These results are in line with the type of answer that questions for facts promote, as will be seen in **Table 7.4**. Although the majority of students' responses in **Extract 7.1** are *one-phrase T-units*, there are also a couple of occasions in which students produce more complex responses, for example, in turn 37 ("*are invertebrates*"), and in turn 16 ("*I don't know what is in English*"). The other students' interventions realized through *one-* or *more-than-one-clause T-units* are those instances in which students use the L1 to try to explain what they mean (see turns 18, 20 and 28).

Extract 7.2 is the beginning of a Citizenship lesson from an AfL group where teacher and students are talking about being healthy.

Extract 7.2 AfL-2, Citizenship, second unit, classes 1 and 2

1	TCH:	What are we going to look at today in Citizenship? Adriana?
2	STU1:	About the %x…x%
3	TCH:	About…excuse me?
4	STU1:	About the health
5	TCH:	About being healthy, good. Um, why are we looking at being healthy? Marcos?
6	STU2:	To have a good diet and…to be good.
7	TCH:	((correction))To be good, to feel good. Good, why else do you think, Bea?
8	STU3:	to know how eh, we are eh, how is the level of our health.
9	TCH:	Good, to know about ourselves and our own health, good, that's a good suggestion, Laura?
10	STU4:	To make our longer life
11	TCH:	To make our lives longer, brilliant.
12		Um, who can remember who…what happened about two or three weeks ago? You had a special visitor to come in and talk to you.
13		Do you remember that?
14	STU:	Yes
15	TCH:	%x…x% people it's always things. Laura, can you tell me about that please?
16	STU4:	Eh, a woman, a woman come to six, to year six, for the blackboard
17	TCH:	Yeah, the whiteboard, whiteboard ((correction))
18	STU4:	Whiteboard. And they and she eh, talked to us about how we can do our life longer like %x…x% says and to have a healthy diet and and to don't em, maybe eh %L1…fumar...L1% ((mimes smoking))
19	STU:	Smoke
20	TCH:	No smoking
21	STU4:	Or eh, for being better than %x…x%
22	STU:	And don't drink! ((interrupting))
23	TCH:	And don't drink and put your hand up as well, yes? Yes, ok, Andrés?
24	STU5:	And they talk about don't eat a lot of meat
25	TCH:	Not eating a lot of meat, ok, why do you think that we should not eat a lot of meat?
26	STU5:	Eh, because, it's bad for the heart ((pronounced incorrectly))
27	TCH:	Bad for your health, yeah, who can expand on that? Who can tell me a bit more about that? Antonio?
28	STU6:	For to not have %L1…grasa…L1%

In contrast with **Extract 7.1**, in this extract only two questions for facts are used (turns 1 and 13). The other questions are questions for reasons (turn 5), questions for explanations (turns 15, 23), and meta-cognitive questions (turns 7, 9, 25, 27). Thus, this teacher's repertoire of questions is more varied than Teacher 3's (see above). This has an effect on students' responses, as most of them are *one-clause* ("*To make our longer life*") or *more than one-clause responses* ("*And they and she eh, talked to us about how we can do our life longer like %x…x% says and to have a healthy diet and and to don't em, maybe eh %L1…fumar...L1%*"). The only examples of *minimal responses* ("*yes*") or *one-phrase responses* ("*about the health*") are turns 2, 4 and 14. The types of questions asked by the Non-AfL teacher in **Extract 7.1** and the AfL teacher in **Extract 7.2** are different, and this seems to have an effect on the complexity of students' responses.

The next piece of results is concerned with the comparison of students' responses between the two groups in each programme. As already specified, it is interesting to see whether significant differences between groups belonging to the same type of programme appear. If the groups in each programme prove to be homogeneous, significant differences between programmes will have a stronger validity. **Table 7.2** below shows the complexity of students' responses in the two AfL groups.

		AfL-1		AfL-2		
	%	N	%	N	*T*	χ^2
Minimal response	16.4%	454	15.7%	90	0.42	0.18
T-Unit response	78.7%	2176	81.2%	465	1.32	1.73
Truncated response	4.9%	135	3.1%	18	1.81*	3.29*
Total		2765		573		
T-Unit						
one-phrase	53.0%	1141	33.6%	155	7.62***	56.87***
one-clause	35.9%	774	31.7%	146	1.74*	3.03*
more than one-clause	11.1%	239	34.7%	160	13.21***	163.73***
Total		2154		461		
Note: *$p < 0.1$; ***$p < 0.001$						

Table 7.2 Students' responses in AfL schools.

Significant differences appear, especially regarding *one-phrase responses*, which are more frequent in AfL-1, and *more-than-one-clause T-unit responses*, which are significantly more frequent in AfL-2. Weaker differences appear in the case of *truncated responses* and *one-clause responses*, both appearing more frequently in AfL-1 lessons. These results seem to indicate that students' responses tend to be more complex in AfL-2, probably influenced by the fact that Teacher 2 asks a lot of meta-cognitive questions and questions for reasons.

In **Table 7.3**, responses from students in the two Non-AfL schools are compared. There are significant differences in all types of responses except for *truncated answers*. *Minimal*

responses appear more frequently in Non-AfL-4 whereas *T-unit responses* are more numerous in Non-AfL-3. However, in Non-AfL-3 most of the *T-unit responses* are *one-phrase* while in the case of Non-AfL-4, *one-clause* and *more than one-clause* responses are more abundant.

	Non-AfL-3		Non-AfL-4			
	%	N	%	N	*T*	χ^2
Minimal response	12.8%	161	19.6%	175	4.27***	18.08***
T-Unit response	82.8%	1040	77.1%	689	3.31***	10.90***
Truncated response	4.4%	55	3.4%	30	1.20	1.44
Total		1256		894		
T-Unit						
one-phrase	71.7%	742	44.1%	300	11.93***	131.55***
one-clause	22.6%	234	39.6%	270	7.71***	57.48***
more than one-clause	5.7%	59	16.3%	111	7.30***	51.70***
Total		1035		681		
Note: ***$p < 0.001$						

Table 7.3 Students' responses in Non-AfL schools.

In this section, both quantitative as well as qualitative data have illustrated the complexity of students' responses in the different groups: significant differences were found out when AfL groups were compared to Non-AfL groups, the AfL groups producing more complex responses. Some significant differences also appeared between AfL-1 and AfL-2, and between Non-AfL-3 and Non-AfL 4. These differences could be related to the type of question asked by the teacher, as will be shown in the next section.

7.2 STUDENTS' RESPONSES TO DIFFERENT QUESTION TYPES

This section explores the relationship between the teacher's type of question and the complexity of students' responses. **Table 7.4** shows the raw numbers and percentages of the types of responses that each type of question triggered. These results are general, which means that no differentiation depending on teachers or subjects has been made. The results show that questions for facts and language questions are the two types generating simpler and shorter responses: 21% of responses to questions for facts are *minimal responses* and, when they generate T-unit responses, it is mainly *one-phrase T-units* (46.5%); language questions also promote 20.9% of *minimal responses* and, when generating a T-unit response, *one-phrase T-units* are also the most frequent ones (56.1%). Leaving aside questions for facts and language questions, this is the order of question types triggering from more to less complex responses: meta-cognitive questions, questions for reasons, questions for explanations, questions for opinions, and meta-questions. It is interesting to remark that questions for reasons are the question type that gets, by far, more *truncated responses* (21.6%).

	Facts		Explanations		Reasons		Opinions		Meta-cognitive		Meta-questions		Language	
	%	N	%	N	%	N	%	N	%	N	%	N	%	N
Minimal response	21.0%	318	4.6%	14	—	—	24.7%	60	2.7%	2	29.7%	95	20.9%	58
T-Unit response	73.6%	1114	88.5%	270	78.4%	69	70.4%	171	92.0%	69	65.9%	211	75.2%	209
Truncated response	3.8%	57	6.2%	19	21.6%	19	3.7%	9	5.3%	4	3.8%	12	2.9%	8
T-Unit														
one-phrase	46.5%	704	16.1%	49	9.1%	8	28.0%	68	—	—	28.7%	92	56.1%	156
one-clause	22.4%	339	43.6%	133	53.4%	47	25.9%	63	37.3%	28	26.9%	86	13.3%	37
more than one-clause	4.3%	65	27.9%	85	15.9%	14	15.6%	38	54.7%	41	9.7%	31	5.8%	16

Note: ***$p < 0.001$

Table 7.4 Complexity of students' responses depending on question type.

Extracts 7.3 and 7.4 below show a prototypical response to a question for fact and a language question, respectively, a *one-phrase response* ("*in the water*" and "*fins?*").

Extract 7.3 Response to question for facts/definition (Non-AfL-3, Science, first unit, fourth class)

1 TCH: Where do they live? Where do we find fish? Nata- eh.. Yasmina
2 STU: in the water

Extract 7.4 Response to language question (Non-AfL-3, Science, first unit, fourth class)

1 TCH: And how do we say <L1 aletas L1> in English, Noelia?
2 STU: fins?

Since language questions and questions for facts foster very similar responses, and since it seems that they are by far the types generating less complex answers, I have taken one of them as the reference (specifically, questions for facts, as they are content questions, like most of the other question types) and have compared the types of responses that questions for facts receive with the types of responses that the rest of question types generate. **Table 7.5** below shows that questions for explanations trigger more complex responses than questions for facts.

	Facts		Explanations			
	%	N	%	N	*T*	χ^2
Minimal response	21.4%	318	4.6%	14	6.92***	46.72***
T-Unit response	74.8%	1114	89.1%	270	5.45***	29.25***
Truncated response	3.8%	57	6.3%	19	1.92*	3.70*
Total		1489		303		
T-Unit						
one-phrase	63.5%	704	18.4%	49	14.26***	177.33***
one-clause	30.6%	339	49.8%	133	6.01***	35.25***
more than one-clause	5.9%	65	31.8%	85	12.93***	149.29***
Total		1108		267		
Note: *$p < 0.1$; ***$p < 0.001$						

Table 7.5 Comparison of responses obtained by questions for facts and questions for explanations.

To illustrate these differences, it is interesting to compare **Extracts 7.3 or 7.4** above with **Extract 7.5** below, in which a question looking for an explanation triggers a *one-clause response* (line 2).

Extract 7.5 Response to question for explanation (Teacher 3, Science, second unit, class 1)

1 TCH: How did they eat?
2 STU: They hunt animals

The same occurs when the responses to questions for facts are compared to the responses to questions for reasons (**Table 7.6**), which trigger significantly more complex responses (with more *one-clause* and *more more-than one-clause responses*).

	Facts		Reasons			
	%	N	%	N	*T*	χ^2
Minimal response	21.4%	318	—	—	0.00	23.54***
T-Unit response	74.8%	1114	78.4%	69	0.76	0.57
Truncated response	3.8%	57	21.6%	19	7.70***	57.15***
Total		1489		88		
T-Unit						
one-phrase	63.5%	704	11.6%	8	8.84***	73.33***
one-clause	30.6%	339	68.1%	47	6.55***	41.49***
more than one-clause	5.9%	65	20.3%	14	4.68***	21.58***
Total		1108		69		
Note: ***$p < 0.001$						

Table 7.6 Comparison of responses obtained by questions for facts and questions for reasons.

In **Extract 7.6**, the teacher is asking a question for reasons. This question triggers a student response consisting of two clauses (two *T-units* really). If we compare this response with the responses in **Extracts 7.3 and 7.4** above, differences in complexity of the student response are clear.

Extract 7.6 Response to question for reason (Teacher 2, Citizenship, second unit, classes 1 and 2)

1 TCH: Why is it not so healthy to eat ice cream? Bea?
2 STU: Because it's like a bomb to your stomach. Because it's so, em, cold

Responses to questions for facts and questions for opinions are compared in **Table 7.7**. Some significant differences appear as well. The number of *one-clause* and *more than one-clause responses* is significantly higher in students' responses to questions for opinions.

	Facts		Opinions			
	%	N	%	N	*T*	χ^2
Minimal response	21.4%	318	25.0%	60	1.27	1.61
T-Unit response	74.8%	1114	71.2%	171	1.17	1.38
Truncated response	3.8%	57	3.8%	9	0.06	0.00
Total		1489		240		
T-Unit						
one-phrase	63.5%	704	40.2%	68	5.84***	33.30***
one-clause	30.6%	339	37.3%	63	1.74*	3.04*
more than one-clause	5.9%	65	22.5%	38	7.55***	54.61***
Total		1108		169		
Note: *$p < 0.1$; ***$p < 0.001$						

Table 7.7 Comparison of responses obtained by questions for facts and questions for opinions.

Extract 7.7 illustrates a question for opinion triggering a *one-clause T-unit response*. Again, the responses triggered by questions for opinions are normally more complex than those triggered by questions for facts or language questions.

Extract 7.7 Response to question for opinion (Teacher 4, Citizenship, second unit, class 3)

1 TCH: What do you think about teachers, Inés? In general, what do you think about them?
2 STU: That there are very good

Meta-cognitive questions trigger the most complex responses, so it is not surprising that differences between their responses and responses to questions for facts are statistically significant (**Table 7.8**). The only similar type of response for the two question types is *truncated responses*, with low percentages in both cases and no significant differences. The percentage of more than *one-clause responses* to meta-cognitive questions in relation the same type of response to questions for facts is more than ten times higher (59.4% as opposed to 5.9%).

	Facts		Meta-cognitive			
	%	N	%	N	*T*	χ^2
Minimal response	21.4%	318	2.7%	2	3.93***	15.33***
T-Unit response	74.8%	1114	92.0%	69	3.39***	11.44***
Truncated response	3.8%	57	5.3%	4	0.66	0.43
Total		1489		75		
T-Unit						
one-phrase	63.5%	704	—	—	0.00	109.09***
one-clause	30.6%	339	40.6%	28	1.74*	3.02***
more than one-clause	5.9%	65	59.4%	41	16.77***	227.33***
Total		1108		69		
Note: **$p < 0.05$; ***$p < 0.001$						

Table 7.8 Comparison of responses obtained by questions for facts and meta-cognitive questions

As a response to a meta-cognitive question, the student in **Extract 7.8** produces a quite complex and long answer (*more than one-clause T-unit*).

Extract 7.8 Response to meta-cognitive question (Teacher 2, Citizenship, second unit, classes 1 and 2)

1 TCH: [...]why else do you think, Bea?
2 STU: to know how eh, we are eh, how is the level of our health.

Finally, regarding the differences between questions for facts and meta-questions, except for *truncated responses*, all the other types of students' answers are significantly different (see **Table 7.9** below). Meta-questions trigger more *minimal responses* but also more *one-clause* and *more than one-clause responses*; questions for facts, on the other hand, foster more *T-unit responses*, but they are mostly *one-phrase*.

	Facts		Meta-questions			
	%	N	%	N	*T*	χ^2
Minimal response	21.4%	318	29.9%	95	3.29***	10.78***
T-Unit response	74.8%	1114	66.4%	211	3.10***	9.60***
Truncated response	3.8%	57	3.8%	12	0.05	0.00
Total		1489		318		
T-Unit						
one-phrase	63.5%	704	44.0%	92	5.35***	28.02***
one-clause	30.6%	339	41.1%	86	3.00***	8.96***
more than one-clause	5.9%	65	14.8%	31	4.61***	20.92***
Total		1108		209		
Note: ***$p < 0.001$						

Table 7.9 Comparison of responses obtained by questions for facts and meta-questions.

Extract 7.9 shows a meta-question followed by a *one-clause student response*. One-clause student responses are frequently triggered by meta-questions, as can be seen in **Table 7.9** above.

Extract 7.9 Response to meta-question (Teacher 1, Science, second unit, class 3)

1 TCH: David, what about your paper? What did you do well?
2 STU: I don't do well

To sum up, certain types of questions seem to trigger significantly more complex responses than others. On the one hand, questions for facts and language questions are the two types that elicit the simplest answers. On the other hand, questions for reasons and meta-cognitive questions foster the most complex responses. In between, triggering more complex answers than questions for facts but less complex answers than meta-cognitive questions, we find questions for explanations, for opinions and meta-questions.

7.3 STUDENTS' RESPONSES ACROSS SUBJECTS

This section will explore the possible effect of the subject on students' responses. Students' responses from each group will be compared in the two different subjects and the complexity of responses to the same question types in the different subjects will also be analysed.

In AfL-1 Citizenship lessons, students' responses are more complex than in Science lessons (see **Table 7.10**), with more one-clause and more than one-clause responses.

	Science		Citizenship			
	%	N	%	N	*T*	χ^2
Minimal response	16.1%	264	16.2%	190	0.07	0.01
T-Unit response	77.5%	1270	77.3%	906	0.11	0.01
Truncated response	5.0%	82	4.5%	53	0.59	0.35
Total		1639		1172		
T-Unit						
one-phrase	43.4%	712	36.6%	429	3.65***	13.25***
one-clause	25.7%	421	30.1%	353	2.60***	6.73***
more than one-clause	7.7%	127	9.6%	112	1.69*	2.87*
Total		1639		1172		
Note: *$p < 0.1$; ***$p < 0.001$						

Table 7.10 Students' responses in Teacher 1's classes of Science and Citizenship.

This result has to be interpreted cautiously, as students' language complexity in this subject may be related to the fact that Teacher 1 asks more questions for explanations in Citizenship than in Science, and more questions for facts and language questions in Science than in Citizenship. However, if we look at possible differences in the complexity of AfL-1 students' responses to the same question types in Science and Citizenship, there are significant differences (illustrated in **Tables 7.11-7.14**).

	Science		Citizenship			
	%	N	%	N	*T*	χ^2
Minimal response	6.4%	17	2.3%	3	1.71*	2.91*
T-Unit one-clause	47.2%	110	59.6%	68	2.19**	4.74**
Total		267		128		
Note: **$p < 0.1$; **$p < 0.05$						

Table 7.11 Students' responses to questions for explanations in Science and Citizenship in AfL-1.

	Science		Citizenship			
	%	N	%	N	*T*	χ^2
Minimal response	26.6%	21	12.5%	15	2.55***	6.38***
T-Unit response	69.6%	55	85.8%	103	2.81***	7.66***
Total		79		120		
T-Unit						
one-phrase	66.0%	35	39.2%	40	3.26	10.05
one-clause	18.9%	10	41.2%	42	2.84	7.79
Total		53		102		
Note: ***$p < 0.001$						

Table 7.12 Students' responses to questions for opinions in Science and Citizenship in AfL-1.

	Science		Citizenship			
	%	N	%	N	*T*	χ^2
Minimal response	11.4%	29	17.7%	23	1.72*	2.94*
Total		255		130		
Note: *$p < 0.1$						

Table 7.13 Students' responses to language questions in Science and Citizenship in AfL-1.

	AfL		Non-AfL			
	%	N	%	N	*T*	χ^2
Minimal response	20.7%	6	5.3%	2	1.96*	3.72*
Total		29		38		
Note: *$p < 0.1$						

Table 7.14 Students' responses to meta-cognitive questions in Science and Citizenship in AfL-1.

Specifically in AfL-1, language questions trigger more *minimal responses* in Citizenship; questions for explanations generate more *minimal responses* in Science and more *one-clause responses* in Citizenship; questions for opinions generate more complex responses in Citizenship; and finally, meta-cognitive questions trigger more *minimal responses* in Science. Therefore, it could be said that students' responses are more complex in Citizenship than in Science not only because of the question types but also because of the subject. As shown in Chapter 6, Citizenship offers space for students to explore their own opinions and reflections. **Extracts 7.10 and 7.11** illustrate a *one-clause response* to a question for opinion in a Citizenship lesson, and a *one-phrase response* to a question for opinion in a Science lesson, respectively, in order to show the different complexity of answers that this type of question triggers in different subjects.

Extract 7.10 Question for opinion, one-clause response. Citizenship

1 TCH: Another opinion could be?
2 STU: <u>You colour very nice</u>

Extract 7.11 AfL-1. Question for opinion, one-phrase response. Science

1 TCH: What does it make you think of? ((referring to a song she's played))
2 STU: <u>A mystery</u>

Responses from students in AfL-2 classes of Citizenship and Drama are also significantly different (see **Table 7.15** below). *Minimal responses* and *more than one-clause responses* are more frequent in Citizenship classes, whereas in Drama classes, there are more *one-phrase* and *one-clause responses*. In relation to the types of questions asked by the teacher in the two subjects, responses would have been expected to be more complex in Drama classes, because in them, Teacher 2 asks much more meta-cognitive questions and meta-questions. However, questions for explanations and for reasons (more frequent in Citizenship classes) also trigger quite complex responses.

	Citizenship		Drama			
	%	N	%	N	*T*	χ^2
Minimal response	16.9%	77	10.2%	13	1.84	3.36
T-Unit response	78.3%	357	85.0%	108	1.68	2.80
Truncated response	2.9%	13	3.9%	5	0.62	0.39
Total		456		127		
T-Unit						
one-phrase	24.3%	111	34.6%	44	2.33**	5.40**
one-clause	22.4%	102	34.6%	44	2.84***	7.98***
more than one-clause	30.7%	140	15.7%	20	3.37***	11.16***
Total		456		127		
Note: **$p < 0.05$; ***$p < 0.001$						

Table 7.15 Students' responses in Teacher 2's classes of Citizenship and Drama.

The comparison of responses to the same type of question in different subjects shows that there are significant differences in the responses to certain types of questions across subjects. In most of the cases, the students' responses given in Citizenship lessons were more complex than the ones given to the same question types in Drama. **Tables 7.16-7.19** illustrate the significant differences that appear when students' responses to each question type are compared in both subjects.

	Citizenship		Drama			
	%	N	%	N	*T*	χ^2
Minimal response	24.6%	35	5.3%	2	2.66***	6.90***
T-Unit response	71.8%	102	94.7%	36	3.02***	8.79***
Truncated response		142		38		
Note: ***$p < 0.001$						

Table 7.16 Students' responses to questions for facts in Citizenship and Drama in AfL-2.

	Citizenship		Drama			
	%	N	%	N	*T*	χ^2
Truncated response	3.8%	3	14.3%	3	1.81*	3.24*
Total		79		21		
T-Unit						
one-phrase	35.4%	23	70.6%	12	2.70***	6.83***
more than one-clause	33.8%	22	11.8%	2	1.79*	3.17*
Total		65		17		
Note: *$p < 0.1$; ***$p < 0.001$						

Table 7.17 Students' responses to questions for opinions in Citizenship and Drama in AfL-2.

	Citizenship		Drama			
	%	N	%	N	*T*	χ^2
Minimal response	1.9%	1	12.5%	4	2.01**	3.96**
T-Unit response	96.2%	50	84.4%	27	1.92*	3.60*
Total		52		32		
T-Unit						
one-clause	32.7%	16	59.3%	16	2.30**	5.06**
more than one-clause	57.1%	28	29.6%	8	2.35**	5.29**
Total		49		27		
Note: *$p < 0.1$; **$p < 0.05$						

Table 7.18 Students' responses to meta-cognitive questions in Citizenship and Drama in AfL-2.

	Citizenship		Drama			
	%	N	%	N	*T*	χ^2
Minimal response	53.7%	22	19.4%	6	3.11***	8.74***
T-Unit response	43.9%	18	77.4%	24	2.99***	8.16***
Total		41		31		
Note: ***$p < 0.001$						

Table 7.19 Students' responses to meta-questions in Citizenship and Drama in AfL-2.

Significant differences appear in four question types: questions for facts, triggering more *minimal responses* in Citizenship, and more *T-units* in Drama; questions for opinions, fostering more *truncated* and *one-phrase responses* in Drama, and more *more-than-one-clause* in Citizenship; meta-cognitive questions, generating more *minimal* and *one-clause responses* in Drama, and more *more-than-one-clause* in Citizenship; and meta-questions, triggering more *minimal responses* in Citizenship, and more T-*unit responses* in Drama. Although in this case differences are not as clear as in the case of AfL-1, Citizenship encourages more *more-than-one-clause responses* than Drama. As an illustration of different complexities of students' responses to the same question type in the different subjects in AfL-2, see **Extracts 7.12 and 7.13**.

Extract 7.12 AfL-2. Meta-cognitive question, more-than-one-clause response. Citizenship

1 TCH: Why you have coloured it in that colour
2 STU: I colour it yellow because yellow is a colour with a lot of energy and when you are laughing you have a lot of energy

Extract 7.13 AfL-2. Meta-cognitive question, minimal response. Drama

1 TCH: Why did you think Andrés was good? Because he was moving?
2 STU: Yes

Extract 7.12 illustrates a *more-than-one-clause response* given to a meta-cognitive question in a Citizenship lesson. **Extract 7.13** shows a *minimal response* to the same type of question (meta-cognitive) in a Drama lesson.

In the case of Non-AfL-3 classes, students' responses are significantly different in Science and in Arts (see **Table 7.20** below). In Science, students construct more complex responses. In this case, it can be claimed that it is the subject and not the question type that plays a major role, since the types of questions the teacher asks are very similar regardless of the subject she is teaching (see Chapter 6).

	Science		Arts			
	%	N	%	N	*T*	χ^2
Minimal response	10.1%	98	21.4%	63	5.16***	26.14***
T-Unit response	85.0%	829	71.5%	211	5.33***	27.83***
Truncated response	4.1%	40	5.1%	15	0.73	0.53
Total		975		295		
T-Unit						
one-phrase	57.7%	563	60.7%	179	0.90	0.80
one-clause	21.6%	211	7.8%	23	5.43***	28.88***
more than one-clause	5.4%	53	2.0%	6	2.44***	5.92***
Total		975		295		
Note: ***$p < 0.001$						

Table 7.20 Students' responses in Teacher 3's classes of Science and Arts

Extracts 7.14 and 7.15 below exemplify the same type of question triggering different types of responses in the two different subjects in Non-AfL-3.

Extract 7.14 Non-AfL-3. Question for fact, minimal response. Arts

1 TCH: So here we have two paintings, yes or no?
2 STU: Yes

Extract 7.15 Non-AfL-3. Question for fact, one-clause response. Science

1 TCH: Give me a characteristic of carnivores
2 STU: They hunt for food

Extract 7.14 instantiates a question for fact triggering a *one-clause response* in a Science lesson and **Extract 7.15** exemplifies the same question type, question for fact, triggering a *minimal response*.

As far as teacher Non-AfL-4 is concerned, **Table 7.21** below shows there are no significant differences between students' responses in his Citizenship classes and in his Arts classes, with the exception of *one-phrase responses*, which are more frequent in Arts classes. Although the difference is not significant, the tendency is for students' answers to be more complex in the Citizenship lessons. This, again, may be related to this teacher asking fewer questions for facts and more questions for opinions and meta-cognitive questions in Citizenship.

	Citizenship		Arts			
	%	N	%	N	*T*	χ^2
Minimal response	19.0%	143	21.2%	32	0.63	0.40
T-Unit response	76.4%	576	74.8%	113	0.41	0.17
Truncated response	3.3%	25	3.3%	5	0.00	0.00
Total		754		151		
T-Unit						
one-phrase	31.7%	239	40.4%	61	2.08**	4.30**
one-clause	30.8%	232	25.2%	38	1.37	1.89
more than one-clause	12.9%	97	9.3%	14	1.23	1.51
Total		754		151		
Note: **$p < 0.05$						

Table 7.21 Students' responses in Teacher 4's classes of Citizenship and Arts.

Tables 7.22-7.25 below show that the same question types trigger more complex responses in Citizenship lessons than in Arts.

	Citizenship		Arts			
	%	N	%	N	*T*	χ^2
T-Unit						
one-clause	31.8%	14	66.7%	4	1.68*	2.78*
more than one-clause	61.4%	27	16.7%	1	2.12**	4.28**
Total		44		6		
Note: *$p < 0.1$; **$p < 0.05$						

Table 7.22 Students' responses to meta-cognitive questions in Citizenship and Arts in Non-AfL-4.

	Citizenship		Arts			
	%	N	%	N	*T*	χ^2
T-Unit						
one-phrase	58.0%	40	100.0%	6	2.06**	4.11**
one-clause	26.1%	18	—	—	0.00	2.06
more than one-clause	15.9%	11	—	—	0.00	1.12
Total		69		6		
Note: **$p < 0.05$						

Table 7.23 Students' responses to language questions in Citizenship and Arts in Non-AfL-4.

	Citizenship		Arts			
	%	N	%	N	*T*	χ^2
Minimal response	—	—	100.0%	1	0.00	3.00*
T-Unit response	100.0%	2	—	—	0.00	3.00*
Total		2		1		
Note: *$p < 0.1$						

Table 7.24 Students' responses to meta-questions in Citizenship and Arts in Non-AfL-4.

	Citizenship		Arts			
	%	N	%	N	*T*	χ^2
T-Unit						
one-phrase	47.7%	146	63.7%	51	2.57***	6.53***
one-clause	44.4%	136	25.0%	20	3.19***	9.96***
Total		306		80		
Note: **$p < 0.05$; ***$p < 0.001$						

Table 7.25 Students' responses to questions for facts in Citizenship and Arts in Non-AfL-4.

As illustrated in the tables above, students' responses tend to be more complex in Citizenship lessons than in Arts lessons. The same question types trigger different students' responses depending on the subject. Meta-cognitive questions trigger more *more-than-one-clause answers* in Citizenship; language questions trigger more *one-phrase responses* in Arts; meta-questions foster more *T-unit responses* in Citizenship, and more mini*mal responses* in Arts; finally, questions for facts generate more *one-clause responses* in Citizenship, and more *one-phrase* ones in Arts. In **Extracts 7.16 and 7.17**, there are illustrations of the same type of question (specifically, a language question) fostering a more complex response in a Citizenship lesson than in a lesson of Arts.

Extract 7.16 Non-AfL-4. Language question, more-than-one-clause response. Citizenship

1 TCH: What is summarize?
2 STU: It's like when you do like <x…x> a story of a story

Extract 7.17 Non-AfL-4. Language question, one-phrase response. Arts

1 TCH: What's another word we know for strange?
2 STU: Weird

To sum up, this section has investigated whether the type of subject influences the complexity of students' responses. The findings show that this seems to be the case in the four groups analysed. Although the frequencies of each question type seem to play a role as well, in some cases the same question type triggers more complex responses in one subject than in the other. In this way, students' responses are more complex in Citizenship lessons in the case of AfL-1, AfL-2, and Non-AfL-4, and in Science lessons in the case of Non-AfL-3. There is no data on Citizenship for Non-AfL-3, which could be the reason for it to be the only case in which students' responses are more complex in a different subject.

7.4 STUDENTS' INITIATING TURNS

Many researchers have shown that it is the teacher who normally initiates discourse in classrooms, and that students have very few opportunities of doing so (van Lier 2000; Lemke 1990; Barnes 1975; Cazden 1988, to mention just a few). In the present study, students' initiating turns have also been analysed, as a way of investigating whether primary school students in CLIL contexts are able and allowed to initiate discourse to make questions or to pursue their own ideas. **Table 7.26** below shows students' percentages of initiating turns across

different subjects. As the table reveals, the greatest number of initiations are found in Arts, followed by Drama. Citizenship and Science are the subjects in which students initiate discourse least frequently. This might be related to the fact that both Arts and Drama are more hands-on subjects, in which students are more frequently asked to work in groups and to carry out projects and more practical activities.

	Science	Citizenship	Arts	Drama
Student instantiations	12.8%	12.4%	28.3%	20.1%

Table 7.26 Students' initiation across subjects.

Some examples of students' initiating turns are illustrated in **Extract 7.18** below. In this extract, two parts can be differentiated. The first part goes from turns 1 to 3, in which there is a prototypical teacher-student interaction, with the teacher asking questions (in this case about a book they are reading). The second part starts in turn 4 with a student initiation.

Extract 7.18 Teacher 1, Citizenship, second unit, first class

1	TCH:	He made a crown ((correction)) very good, Luis, and now what is he going to do with the branches? Yoohoo!! ((mimes swinging))
	[…]	[…]
2	STU:	Swing
3	TCH:	Swing, he's going to swing from her branches.
4	STU1:	Therese, but I say the crown and don't Luis, I said it. (student initiation)
5	TCH:	It was Andrea but I saw Luis's
6	STU:	I don't say it
7	TCH:	I saw your mouth say ((mouths crown))
8	STU1:	No!
9	TCH:	Did you say crown?
10	STU2:	yes
11	STU:	Yes

The first student initiation in **Extract 7.18** is related to the topic at hand, but the student wants to make clear it was her who gave the answer first, and not another student as the teacher thinks (she is, then, stating a fact). This student initiating turn gives way to a series of turns in which what the student has said in her initiation is discussed.

Many times, though, student initiations are not followed by such an extended exchange. As student initiations are usually questions, these are followed by the teacher´s response, but immediately the topic and pace of the class is recovered, with the teacher playing the role of the initiator again. This is illustrated in **Extract 7.19** below.

Extract 7.19 Teacher 3, Science, second unit, first class

1 TCH: Ok? Bison was another animal they hunted in those times
2 STU1: what is bison? (student initiation)
3 TCH: Ok, bison, <L1 bisonte L1>
4 STU1: ah, <L1 bisonte L1>
5 TCH: So here you have some characteristics [that I told you]

In **Extract 7.19**, a student does not know the meaning of "bison", and so he initiates an exchange with a question (turn 2). After the teacher has answered the question, the normal flow of the lesson is soon re-established (turn 5).

Table 7.27 below illustrates all students' initiations now classified into groups and subjects within each group. Percentages have been calculated out of the total number of initiations.

	AfL-1	AfL-2	Non-AfL-3	Non-AfL-4
Science	128 (7.2%)		256 (20.7%)	
Citizenship	129 (9.9%)	37 (7.5%)		172 (18.5%)
Arts			136 (31.5%)	40 (20.9%)
Drama		32 (20.1%)		
Total	257	69	392	212

Table 7.27 Students' initiations across subjects and schools.

Looking at the percentages, results indicate that it is in Non-AfL lessons where the students make more initiating turns. This might indicate that using an AfL methodology does not stimulate the frequency of students' initiating turns.

After having analysed students' initiations in each group, let us now focus on the two types of programme. Comparisons between the two groups of each type of programme have not been included because there were no significant differences. Therefore, in this case, the results in AfL and Non-AfL schools were homogeneous. **Table 7.28** displays the comparison of students' initiations in AfL schools and in Non-AfL schools.

	AfL		Non-AfL			
	%	N	%	N	*T*	χ^2
Question	30.1%	98	57.0%	344	8.10***	61.40***
content	67.3%	66	67.1%	230	0.05	0.00
language	20.4%	20	26.2%	90	1.18	1.38
content+language	12.2%	12	6.7%	23	1.79*	3.20*
State fact	32.2%	105	19.4%	117	4.42***	19.20***
Explaining	13.2%	43	7.9%	48	2.57***	6.59***
Personal opinion	18.4%	60	9.3%	56	4.05***	16.18***
Argue	5%	16	0.7%	4	2.31***	5.34***
Personal experience	1.2%	4	5.8%	35	3.33***	10.99***
Total		326		604		

Note: $^{**}p < 0.05$; $^{***}p < 0.001$

Table 7.28 Students' initiations: comparison between AfL and Non-AfL schools.

Although there are no significant differences regarding the number of initiating turns when programmes are compared, significant differences do appear regarding the types of initiations. When compared to AfL schools, it is significant that, in Non-AfL schools, a high percentage (57%) of students' initiations are questions (more specifically, 67.1% of all the student initiations in Non-AfL schools are questions related to content). The other types of initiations, although present as well, appear more scarcely (e.g. see percentages of stating facts, explaining, giving personal opinions, relating to personal experiences). Arguing is practically absent in Non-AfL classes, whereas in AfL classes they amount to 5% of these students' initiations. In spite of the fact that students also ask questions in AfL classes, they more frequently contribute to classroom discourse by stating facts (32.2%). Other frequent student initiations in AfL classrooms have to do with explanations and giving personal opinions. Regarding question types in both types of programmes, students tend to ask more questions related to content than related to language. This might be surprising because of the demands that the foreign language places on students. However, manipulating content in a foreign language does not only pose difficulties in the use of the foreign language but also in the understanding of that content. Other times, students' questions were related to both content and language. Students in AfL schools ask significantly more questions in this third category (12.2% as opposed to 6.7% in Non-AfL schools).

Due to its high frequency, it is worth illustrating students' *initiations through questions*. **Extract 7.19** above included an example of a student-initiated exchange with a question related to language ("*what is bison?*"). **Extracts 7.20 and 7.21** below illustrate questions related to content and to both content and language. In **Extract 7.20**, the student makes a question related to content, through which he wants to know what the image on the screen represents.

Extract 7.20 Teacher 3, Science, second unit, first class

1	TCH:	<L1 Sí L1>, like buffalos, or bisons, and they also made sculptures, ok? Like this one ((pointing at the image))
2	STU:	What is that?
3	TCH:	[In that period].. In that period they were eh.. mm... they gave a lot of importance to the fact women, no? could have babies, no? Because if women didn't have babies, then.. they could be also in danger of extinction. So it is believed that that is like a.. little sculpture that was done so that women could continue having children, no? A statue for.. a sculpture for fertility, so that women could reproduce..

In **Extract 7.21**, "*what's a stereotype?*" is concerned with both content and language. It is part of the content they are learning but it is a language vocabulary question, at the same time. This

question is, therefore, not specifically related to the second language. It would also be expected to happen in students' L1.

Extract 7.21 Teacher 4, Citizenship, second unit, third class

1	TCH:	Shhh, Irene, we don' know yet that, OK, but the stereotype, do you think that the stereotype of the Spanish people, do you think is a nice stereotype or not?
2	STU:	%L1...Pues no...L1%
3	STU:	What's a stereotype? (student initiation)
4	TCH:	Well, the word we were speaking now, it's an idea of a group, OK? Do you understand?

The second most frequent student initiation type is *stating a fact* (see **Extract 7.18** above and **Extract 7.22** below).

Extract 7.22 AfL-1. Citizenship, first unit, class 1.

1	TCH:	Apostrophe and S. this is the word …what's….what's…what's the date? Many people write this ((writes word)) this is wrong, this isn't a word in English
2	STU:	This is Spanish (student initiation)

In **Extract 7.22**, the teacher is correcting a word and, in the next turn, one student makes an initiating turn saying "this is Spanish", hence stating a fact.

The third most frequent student initiation type is *giving a personal opinion* (18.4% in AfL schools; 9.3% in Non-AfL schools), instantiated in **Extract 7.23** below.

Extract 7.23 Teacher 3, Arts, second unit, first class

1	TCH:	<L1 Claro L1>, of course, if you, you, you could transform like an everyday object into something different. In here, since Coca-cola was a very popular drink at that time so Andy Warhol, no? took as a subject matter for his painting the Coca-col- the drink, no? But at the same time, he made like an advertisement for the company, ok? And the other person, who's the other person?
2	STU:	((Some)): Michael Jackson
3	TCH:	No, that's not Michael Jackson ((some students laugh))
4	STU:	It's very..
5	TCH:	But he was also a very popular [musician]
6	STU:	[It's] awful (student initiation)

During an activity about Andy Warhol, in turn 6 one student gives her opinion about one of Andy Warhol's paintings ("*it's awful*"). In this case, the initiation is not followed, yet there are

cases of students' initiations through personal opinions that generate a longer exchange (see **Extract 7.24** below).

Extract 7.24 Teacher 4. Citizenship, second unit, class

1	TCH:	Yes, Mohammed, you have the same idea? Good. We are going to read the paragraph number three, OK?, and we're going to summarize it too. OK, so can you... for example, Luisa, you want to read it, paragraph number three. It's the grey one
2	STU1:	I think paragraph three we read it (student initiation)
3	TCH:	Don't think...
4	STU1:	It was of the tertiary graduates
5	TCH:	But we spoke of the tertiary because someone asked
6	STU:	Yes
7	TCH:	Yes, but we didn't read all of it, OK?

Explaining is another type of student initiation that was relatively frequent (13.2% and 7.9% in AfL schools and Non-AfL schools, respectively). **Extract 7.25** below includes an example of this type of student initiation.

Extract 7.25 Teacher 2, Citizenship, second unit, classes 1 and 2

1	TCH:	Any questions?
2	STU:	No
3	STU:	No
4	TCH:	Ma…ah, Pedro?
5	STU1:	That my mother told me that being healthy with the %x...x%, but, very, very, very healthy with that becomes unhealthy because my mother told me that when you brush a lot, and then, for example, your teeth, you get out protections and here, for example, little pieces of skin… (student initiation)
6	TCH:	Hmm
7	STU1:	In the shower
8	TCH:	Ok, I see a proper point that Pedro is talking about. Anything that you do too much can become unhealthy. So when I'm saying to you about healthy food, I don't mean for you to stop eating pizza, don't stop eating ice cream, just be careful of what you eat. I eat lots of…I eat pizza and ice cream, %x...x% people. Ok? It's good, it's good to have what you want, anything that you do too much, if you stop eating anything that you think is unhealthy, that would become really unhealthy as well. You just need to have the right balance. Keep clean, but don't scrub away your skin. Yeah? Ok.

The teacher accepts one of the students' bid to intervene, and this student explains, in a very long turn which continues in turn 7, how being too healthy can be unhealthy (turn 5). This student´s contribution is reformulated into a clearer explanation by the teacher (turn 8) through the use of a functional recast (Mohan & Beckett 2003), with the purpose of clarifying the student's argument.

Extract 7.26 below shows an example of a student initiation through *arguing*. This is part of a true/false exercise that students had done at home.

Extract 7.26 Teacher 3, Science, first unit, fifth class

1	TCH:	They have the eyes on the front and not on their sides, that's another characteristic. Is there.. is there anybody who has got a different characteristic? Alba, do you have a different one? Gabriela? Yes?
2	STU1:	eh.. they no have hair.. no.. they have hair
3	TCH:	They have hair. Alba
4	STU2:	One thing. That in the cetaceans (student initiation)
5	TCH:	In cetaceans, yes
6	STU2:	they have no hair is false because for- because for example a seal has hair (student initiation)
7	TCH:	In general. Even dolphins when they are born have got a little bit of hair, yes, you are right. But if we compare a dolphin with a monkey, no?
8	STU2:	yes
9	TCH:	Then, dolphins don't have hair compared to monkeys, no?

In turns 4 and 6, a student initiates a new exchange, arguing why she thinks that the sentence "cetaceans have no hair" is false (in response to another student's claim for this sentence to be true). This argumentation leads the teacher to follow the line of discourse the student started and try to make her see that the sentence is true because they are speaking in general, and so the exchange extends several more turns. In this example, there is a clear cognitive engagement around whether cetaceans have hair or not and the difference it makes whether we are talking in general or not in order to answer the content question.

Extract 7.27 below exemplifies an initiation which includes both *opinion* and *arguing*. In this extract, students are giving each other marks for an activity they have done (peer-assessment). When the mark for one couple has already been decided among teacher and students, then one student makes an initiation explaining why she disagrees with the final decision (turns 2 and 4). The student's initiation pushes the teacher to make further comments and questions (turns 3 and 5), making the student aware of the criteria she has to use in order to assess her classmates.

Extract 7.27 Teacher 1, Citizenship, first unit, second class

1	TCH:	NOTABLE? NOTABLE! NOTABLE, ok, NOTABLE.
2	STU:	I think it's a BIEN because he didn't do gestures and.. (student initiation)
3	TCH:	He did some gestures.
4	STU:	And he, and he do a space. (student initiation)
5	TCH:	But did he do some gestures?
6	STU:	But…

In relation to students' initiations to convey *personal experiences*, the data showed long contributions (see **Extract 7.28**, turn 2), which sometimes went over several turns (see **Extract 7.29**, turns 2, 4, 6, 8). In the case of **Extract 7.28**, the student's initiation triggers the teacher's question in turn 9, and so the exchange extends until turn 11.

Extract 7.28 Teacher 2, Citizenship, second unit, classes 1 and 2

1	TCH:	Who has ever made their own pizza? ((Students raise hands)) Not only is it healthier, but also it's quite fun because you can get the base, the bottom of the pizza, and you can take a little bit of this, and a little bit of that, you can experiment with all the different flavours, and it's much healthier because you can get fresh vegetables and you don't have to put loads of cheese and loads of oil on it, and loads of butter, just make it really nice and fresh. Andrés?
2	STU1:	That with my mother I helped to do a pizza with…well, sometimes we put eh, no healthy, no healthy food but only…is sometimes not all. (student initiation)
3	TCH:	Yeah, you can have a little bit of unhealthy food sometimes, %x...x% you've got to enjoy that as well. Ok, right, let's have a look, what's the next one?

Extract 7.29 Teacher 4, Citizenship, second unit, fifth class

1	TCH:	OK, for example %x...x% Right, %x...x% clear when we are with someone, OK?, is to be happy, yes or no? OK, is the same with bullying, I mean, we have friends, OK?, it's meant that we are going to be happy with our friends, yes or no? OK? Not to fight, right? Good, David.
2	STU1:	That when I was in second year (student initiation)
3	TCH:	Yes
4	STU1:	A boy, a boy of fourth year (student initiation)
5	TCH:	Yes
6	STU1:	He hits me and (student initiation)
7	TCH:	Aha
8	STU1:	And it's bullying me (student initiation)
9	TCH:	OK, he was bullying. Did you tell your teachers?
10	STU1:	No, my mother, and my mother said to the director and
11	TCH:	OK, And you solved the problem

In this section, where student initiating turns have been the focus, results have shown that, as expected, students do not initiate exchanges as often as teachers. Nonetheless, there were some subjects which prompted more student initiations. These subjects were Arts and Drama, probably due to their hands-on nature. Apparently, results also indicate that the methodology (AfL vs. Non-AfL) does not seem to have an effect on the amount of student initiations. However, when the types of initiations were compared in AfL and Non-AfL schools, it was found that there was a greater variety in AfL schools (questions, stating facts, giving personal opinion, explaining, etc.). In turn, in Non-AfL schools, more than half of student initiating turns

were *questions*. Usually, once these questions were answered by the teacher, he/she retook control of the interaction by producing the next initiating turn.

7.5 DISCUSSION

In this chapter, we have presented the results on the second move of IRF patterns: students' responses. This pattern has been criticized for various reasons that will be discussed below, but the main one being that they involve the teacher talking most of the time[4], and most of the teacher's interventions are questions or lecture talk (Mercer 1995). This shows a picture of classroom interaction as being asymmetrical and unequal: the teacher has the right to select the topic and allocate turns (Mehan 1979; Cazden 2001; Sinclair & Brazil 1982), as well as initiate (and end) most IRF exchanges (Lemke 1982, 1990; Nassaji & Wells 2000: 376; Carlsen 1991; Markee & Kasper 2004; Cazden 1986, 2001; van Lier 1996; Mercer et al. 2004; Lemke 1982). Teachers can also address any student at any time, interrupt them, and control their behaviour (Cazden 2001). All this makes students' opportunities to participate in the co-construction of learning scarce (Barnes 1975; Lemke 1982). They are often not allowed to intervene with long and complex turns, they can barely pursue their own ideas instead of the teacher's (van Lier 1996), their creativity and independent thinking are compromised, their interpretations are usually not valid, they are not responsible for their learning, and as a result, teaching effectiveness is jeopardized (Allwright & Bailey 1991; Barnes 1975, 1976, 1982, 1992; Cazden 1988; Gutierrez 1994; Dinsmore 1985; Ellis 1994; Lemke 1982, 1990; Wood 1992; Nystrand 1997; van Lier 2000; Lemke 1990; Hall & Walsh 2002; van Lier 1988, 1996; Hardman et al. 2003).

This chapter has revealed that primary school students are able to produce a wide variety of responses in the L2: from yes/no answers to a whole T-unit with several clauses in it. Teachers' questions have been the object of debate, mainly because they often do not encourage students to produce long responses (Dillon 1990). Since teachers are said to do most of the talking, it is very frequent that students' responses are short and simple (Alexander 2004; Ruiz-Primo & Furtak 2006, 2007; Nassaji & Wells 2000; Nystrand & Gamoran 1997; Hardman et al. 2003; van Lier 1996; Dalton-Puffer 2009). In the case of the present study, results confirm that the majority of students' responses short: 41.98% of students' responses in all classes are *one-phrase responses*; 15.8% are *minimal responses*; 4.27% *truncated responses*. However, the overall percentage of long and, mainly, complex responses, is far from negligible: 35.79%

[4] To my knowledge, the only study in teacher-fronted lessons that contradicts this finding is Dillon (1998), in which students talked as much as the teacher.

(25.57% *one-clause responses*; 10.22% *more-than-one-clause answers*). Thus, our results would be in line with Black et al. (2003), who found contradicting evidence in L1 (English) content classes: in some lessons students contributed with short phrases and in others students formed complete sentences. Likewise, in immersion contexts, there are studies claiming that students are not given many opportunities to talk, with fewer than 15% of their utterances consisting of one clause (Allen et al. 1990), whereas some others found that immersion students were engaged in extended language production (Day and Shapson 1996).

Some factors that have been shown to have an effect on the length and complexity of students' responses are the methodology (Barnes 1975), the type of questions asked by the teacher (Dalton-Puffer 2007; Nassaji & Wells 2000), and the type of subject (Black & Wiliam 2009). This study has proven that there are significant differences in the types of responses given by students depending on the methodology (AfL vs. Non-AfL). When responses in AfL classes and in Non-AfL classes were compared, the results showed that students answered significantly more frequently with *one phrase* in Non-AfL classrooms (47.9% as opposed to 38.2% in AfL classes), whereas students in AfL classes offered more *one-clause* (27.1% as opposed to 23.2% in Non-AfL classes) and *more-than-one-clause responses* (11.8% as opposed to 7.8% in Non-AfL schools). These results verify one of the hypotheses in this study: there are significant differences in the types of responses given by learners in AfL and Non-AfL classes. The fact that students' responses are longer and more complex in AfL classes may lead to the interpretation that CLIL teachers following an AfL methodology do not exercise their control on classroom interaction as tightly as others do. This interpretation is in line with the findings obtained in the previous chapter, where AfL teachers show a less authoritative control of the interaction, as they ask fewer questions for facts and more questions for reasons, for opinions, and meta-cognitive questions. It seems to be the case that individual teacher styles may also play a role, as there were some significant differences across groups in each programme (AfL and Non-AfL).

This study also proves that, depending on the type of question asked, shorter or longer answers are triggered, something which has already been claimed by other researchers (Dalton-Puffer 2007; Nassaji & Wells 2000). Fewer questions for facts and more for opinions and meta-cognitive questions (Dalton-Puffer 2006, 2007) have been seen to lead to a more dialogic type of discourse and a more exploratory stance regarding the topic (Nassaji & Wells 2000: 381; Mercer 1995; Nystrand 1997; Rogoff 1994; see also Dalton-Puffer 2007, 2008). Questions for facts often get short responses, whereas questions for opinions trigger longer contributions (Dalton-Puffer 2008; Mercer et al. 2004). These findings are also confirmed by the results in the present study. Likewise, Harrison and Howard (2009) emphasize that questions which promote

discussion and collaboration require longer answers and prioritize formative opportunities. Nassaji and Wells (2000) found a positive correlation between length and complexity of students' responses and negotiatory questions, and a negative one between length and complexity of the response and factual questions. Again, this is also true in the case of the present study, in which questions for facts usually elicited short, simple answers, whereas questions for opinions, for reasons, and meta-cognitive questions tend to trigger higher complexity. In the same line as previous studies and the present one, Harlen and Winter (2004) claim that, if the focus is not on the correctness of the answer but on what students think and what their conceptions are, responses will not consist of just one word but will be longer and more complex. At the same time, they argue (Harlen & Winter 2004) that this technique fosters deep learning as opposed to rote memorization, in line with AfL pedagogy. In line with Llinares and Morton (2010), who found that students' turns in which explanations were involved were longer than regular turns, in this study students' responses in which they had to provide explanations have also shown a higher complexity.

Although the context of the present study is new and different (involving CLIL and AfL), the results displayed in this chapter do nothing but confirm all the results described above in previous studies: questions for facts generate the shortest and simplest responses (along with language questions), as most of their responses are minimal responses or one-phrase responses. This contrasts with meta-cognitive questions, questions for reasons, questions for explanations, and questions for opinions, which tend to trigger complex student responses. As Mercer et al. (2004) state, the more students have to explain and justify their viewpoints, the longer and more complex their responses will be. Furthermore, they will understand the subject better (Mercer 2000) and, in the case of CLIL, they will also process the L2 more deeply. It is interesting that meta-questions is the type which obtains the highest number of truncated responses (but they also get complex responses when not truncated). This may be explained by the question type: the cognitive demand of this type of question is very high, as meta-questions force students to reflect about their learning and assessment, and so sometimes (primary school) students may need to be helped with this type of reflection, particularly if this is done in a foreign language. In addition, Van Lier (1988: 10) warns that sometimes, the cognitive demand of a question does not involve an equally complex language use (for instance, a meta-question that involves students self-assessing themselves may often involve a one-phrase response). This can be especially expected to be the case in CLIL classrooms, in which students are using a foreign language to learn academic content and their level of the L2 may not match their cognitive level (Coyle et al. 2010). In the same line, Carlsen (1991) claims that high level questions do not always correspond to high level thinking and high level responses. In spite of these claims, this

and previous studies (mentioned above) have proven that, normally, complex questions elicit complex responses, even in the L2. Therefore, it can be claimed that the L2 is not a deterrent for students to construct complex responses.

Students' responses also need to be analysed and interpreted within the context in which they have been produced. Different subjects may have an effect on them (Black & Wiliam 2009). In this investigation, the results indeed show that students' responses to the same teacher vary depending on the subject. In Teacher 1's classes, students produced more complex responses in Citizenship than in Science; in the case of Teacher 2, her Citizenship classes also generated more more-than-one-clause responses whereas in Drama there were more one-clause and one-phrase answers; students in Teacher 3's classes construct more complex responses in Science than in Arts; and, finally, in the case of Teacher 4, learners tend to create more complex responses in Citizenship than in Arts. In order to check whether the question types could also have an influence on these differences, students' responses to the same question types in the different subjects were compared. The findings revealed that, in certain subjects, students' responses to the same questions were more complex than in others. These results confirm one of the starting hypotheses, namely, that there would be differences across subjects and, indeed, we have seen how responses are significantly different depending on the subject.

Finally, regarding students' initiations, previous research has shown that students do not have many chances to initiate discourse and ask their own questions (Cazden 2001; Sinclair & Brazil 1982; van Lier 2000; Carlsen 1991; Rosenshine 1976; Wragg & Brown 2001), and that encouraging students to initiate topics and ask questions can lead to higher engagement and spontaneous participation (Mortimer & Scott 2003), learning more content (Aulls 1998), performing better in tests (Black & Wiliam 1998a), and learning more or richer language (Llinares 2007). Students' initiations represent a way of keeping students active during the lesson, with positive effects on learning (Wragg & Brown 2001). Some researchers claim that there is the need of moving away from IRFs for students to pursue their own ideas (van Lier 1996), whereas others think that small changes in the existing IRFs would have a positive effect (Cazden 2001). In line with the majority of previous studies, the results of the present study reveal that students' initiations, when compared to teachers', are few in both AfL and Non-AfL contexts, in spite of the fact that CLIL contexts have been claimed to generate more students' initiations (Nikula 2007). A total of 930 exchanges are student-initiated, as opposed to 3826 teacher-initiated, which gives an approximate ratio of 1 out of 4. Therefore, using an AfL methodology cannot be claimed to have an effect on increasing the number of students' initiating turns. When initiations are explored by subject, Arts and Drama display the highest percentages of student-initiated exchanges. A possible explanation for this finding might be that

both subjects involve students in group work or individual activities, thus prompting them to initiate more exchanges with the teacher or their classmates. Regarding the types of student initiations, more than half of students' initiations in Non-AfL classes are questions, normally content questions. On the contrary, students' initiations in AfL classes are more varied, which may indicate that these students feel more comfortable with initiating discourse even when they are not asking questions, as AfL teachers do not seem to control interaction as much as Non-AfL teachers.

In sum, AfL classrooms seem to match the characteristics of dialogic teaching more than AfL classrooms as teachers' questions seek more students' opinions and reasoning; students' responses are longer and more complex; students are allowed to initiate discourse, not only to ask questions but also to provide explanations or argue a certain viewpoint.

7.6 SUMMARY AND CONCLUSION

This third chapter on results has completed another move in the IRF exchange. Students' responses have been analysed and compared across AfL and Non-AfL classes This comparison has shown that students' responses are significantly more complex in AfL groups than in Non-AfL ones. Therefore, one of the initial hypotheses is confirmed, that is, there are significant differences in the types of responses given by students in AfL schools and Non-AfL schools. Secondly, students' responses have also been compared in terms of the type of question asked by the teacher, and results have revealed that the question type affects the quality of the response: questions for facts obtain simpler answers whereas meta-cognitive questions, questions for reasons or for opinions trigger more complex responses (in line with previous studies). The subject has also turned out to be important (again, confirming an initial hypothesis), as the complexity of students' responses differ depending on the subject they are in. As for students' initiations, and validating previous research, they are relatively infrequent, especially when compared with teachers' questions. Student-initiated exchanges are not affected by the methodology (AfL vs. Non-AfL), but it seems that the subject does have an effect on their frequency (with Drama and Arts prompting more student initiations than Science and Citizenship). The methodology is important when the types of initiations are compared in AfL and Non-AfL groups. In Non-AfL classes, the majority of student-initiated exchanges are questions. In contrast, in AfL lessons, students' initiations are more varied, since different types are considerably more present. These results further confirm the general tendency that seemed to emerge in the previous chapter: classes in which AfL is implemented appear to be more formative and dialogic than classes in which AfL is not used (Alexander 2004; Black et al.

2003). The present research seems to indicate that formative and dialogic classes in which questions for reasons, for explanations, meta-cognitive questions, and meta-questions are asked engage students both at the cognitive and language level. This is crucial in CLIL contexts, where the integration of content and language is at the fore.

8

Results IV — Teachers' feedback

This chapter will be devoted to the analysis of teachers' feedback across groups (AfL, Non-AfL). The types of feedback used by AfL and Non-AfL teachers will be analysed in a variety of ways: in relation to the group, to the question type, to student responses, and in relation to the subject taught. Finally, a qualitative analysis of the effect of different types of feedback on student uptake will also be offered. As in previous chapters, the results obtained will be discussed at the end of the chapter.

8.1 GENERAL USE OF FEEDBACK TYPES BY ALL TEACHERS IN ALL SUBJECTS

First of all, the use of feedback types used in general by all teachers will be presented. For each feedback type, the mean percentage was calculated (see **Table 8.1**).

Evaluation	Expansion	Revision	Re-route	Meta-feedback
42.6%	25.3%	3.2%	20.1%	8.7%

Table 8.1 General use of feedback types.

As explained in Chapter 4, the types of feedback analysed in this study are *evaluation* (teacher's assessment of the response as correct or incorrect), *expansion* (the teacher pushing students to continue talking by asking further questions or offer related information), *revision* (teacher's revision or summary of concepts), *re-route* (teacher's attempt to guide the students to the correct response after an incorrect one), and *meta-feedback* (teacher's feedback on students' learning and assessment processes). As expected, **Table 8.1** shows *evaluation* is the most frequent type of feedback used, followed, in this order, by *expansion*, *re-route*, *meta-feedback*, and *revision*.

8.2 TEACHERS' FEEDBACK TYPES IN AFL AND NON-AFL CLASSROOMS

After presenting the results on feedback types in general, individual teachers' use of types of feedback will be next put forward (see **Table 8.2** below). Teacher 1, an AfL teacher, uses *evaluation* quite frequently (41.8%), followed by *expansion* (26.2%), *re-route* (17.9%), *meta-feedback* (12.7%), and *revision* (1.4%). Teacher 2, also AfL, uses similar types, 44.7% for *evaluation*, 28.7% for *expansion*, 17% for *re-route*, 8.8% for *meta-feedback*, and 0.8% for *revision*. Teacher 3 is a Non-AfL teacher, and *evaluation* makes up 39.9% of her feedback, *expansion* 23.1%, *re-route* 21.7%, *revision* 9%, and *meta-feedback* 6.3%. Teacher 4, also a Non-AfL teacher, uses 44.4% of *evaluation*, 23.7% of *re-route*, 23.4% of *expansion*, 7.1% of *meta-feedback*, and 1.7% of *revision*.

	AfL			Non-AfL
	Teacher 1	Teacher 2	Teacher 3	Teacher 4
Evaluation	41.8%	44.7%	39.9%	44.1%
Expansion	26.2%	28.7%	23.1%	23.4%
Revision	1.4%	0.8%	9.0%	1.7%
Re-route	17.9%	17.0%	21.7%	23.7%
Meta-feedback	12.7%	8.8%	6.3%	7.1%

Table 8.2 Teachers' feedback types.

Evaluation, the most frequent type of feedback for all teachers, is shown in **Extract 8.1** below.

Extract 8.1 Teacher 1, Citizenship, first unit, class 3

1 TCH: Is this right?
2 STU: Yes
3 TCH: No, it isn't right

Specifically, the kind of *evaluation* found in **Extract 8.1** is *negative evaluation with no explicit correction*. The teacher only negates the previous student response without explaining why it is wrong.

Teacher 2 is the one who uses *expansion* more frequently, an example of which is shown in **Extract 8.2**.

Extract 8.2 Teacher 2, Citizenship, second unit, classes 1 and 2

1 STU1: And they talk about don't eat a lot of meat
2 TCH: Not eating a lot of meat, ok, why do you think that we should not eat a lot of meat?
3 STU1: Eh, because, it's bad for the heart ((pronounced incorrectly))
4 TCH: Bad for your health, yeah, who can expand on that? Who can tell me a bit more about that?
5 STU2: For to not have %L1…grasa…L1%

Extract 8.2 shows how the teacher tries to make students *expand* on their ideas and argue their points through further questions (meta-cognitive questions).

Revision, which is the least frequent type of feedback for all teachers, is illustrated in **Extract 8.3**.

Extract 8.3 Teacher 3, Arts, first unit, classes 1 and 2

1	TCH:	Do you understand the two.. characteristics of cubism?
2	STU:	((Some)) yes
3	TCH:	<u>One is the perspective, ok? And the other is.. a- is the geometric shapes that they use when they represent objects, ok?</u>

In turn 3 from **Extract 8.3**, Teacher 3 is reviewing the characteristics of cubism, which she has just explained earlier in the lesson.

Re-route is exemplified in **Extract 8.4** below. It is an example found in one of Teacher 4's lessons.

Extract 8.4 Teacher 4, Citizenship, second unit, class 1

1	TCH:	And are they slower and less safe than conventional cars?
2	STU:	Yes
3	TCH:	<u>Are you sure? Do you think they are slower?</u>
4	STU:	the same

In this extract the teacher is *re-routing* the student response in turn 2 through a prompt that serves for the student to realize that what she said before was wrong.

The teacher that most frequently used *meta-feedback* was Teacher 1. **Extract 8.5** below illustrates this type.

Extract 8.5 Teacher 1, Citizenship, second unit, class 3

1	TCH:	<u>So my comment is: Dani, when you get a very, very good mark on one paper, you can't go down on the other papers. Right? This is <L1 evaluación continua L1>.</u>

In **Extract 8.5**, Teacher 1 is giving the student feedback related to his learning process and how, in order to keep improving, he should not be satisfied with just one good assignment.

Table 8.3 below shows the results obtained for types of *evaluation* and types of *prompts*. *Evaluation* was further subdivided into *positive* (*with recast* or *with no recast*) or *negative evaluation* (*with explicit correction* or *with no explicit correction*). Both *with recast* and *with explicit correction* were further analysed into focus on *content, language,* or *both*. *Re-route* was also further subdivided into two different types: *recast* and *prompt*, which were further classified according to whether they focused on *language*, *content*, or *both*. The results show that *positive evaluation with no recast* predominates for all teachers, especially for Teacher 2 (95% of evaluation is positive evaluation). An example can be seen in **Extract 8.6** below (turn 3).

Extract 8.6 Teacher 2, Citizenship, second unit, classes 1 and 2

1	TCH:	What makes you become... what puts you in a good mood, what makes you in a good mood, makes you feel happy?
2	STU:	Being with my friends
3	TCH:	Being with your friends.

When there is *positive evaluation with recast*, this tends to focus on *language* in the case of all teachers (92.9% for Teacher 1, 93.2 for Teacher 2, 93% for Teacher 3 and 89.7% for Teacher 4). As an illustration, see **Extract 8.7**, in which the *recast* focusing on *language* is in turn 2: the teacher is recasting "throw the rubbish" into "throw away the rubbish" at the same time as he acknowledges the response as correct.

Extract 8.7 Teacher 4, Citizenship, second unit, class 1

1	STU:	Ahh! Throw the rubbish!
2	TCH:	Throw away the rubbish. Throw ((writes)) away the rubbish, ok? Rubbish. Good. Ok, what else do we have?

Within *negative evaluation*, it is common that an *explicit correction* is included, especially in the cases of AfL Teachers (66.4% for Teacher 1 and 63.6% for Teacher 2). This is exemplified in **Extract 8.8** below, in which Teacher 1 negatively evaluates the student response, but she explicitly includes the correct answer in her turn.

Extract 8.8 Teacher 1, Citizenship, second unit, class 1

1	TCH:	If that's Sergio we have to change it to say?
2	STU:	Her
3	TCH:	He. It's a boy
4	STU:	His

Whether *explicit correction* focuses on content, language, or both varies across teachers: in the case of Teacher 1, she mainly focuses on *language* (56.1%); Teacher 2 mainly focuses on both *content* (42.9%) and *content + language* (42.9%); Teachers 3 and 4 concentrate more on *content* (76.9% and 48.6%, respectively).

Moving on to types of *re-route*, *prompts* are more common for all teachers (Teacher 2 being the one with more balance: 59% *prompts* and 41% *recasts*). **Extract 8.9** below shows two examples of *re-route* in a Science lesson by Teacher 3: in turns 3 and 5, Teacher 3 is trying to re-direct the student's attention to the right picture and, in this way, the student is able to answer the question correctly.

Extract 8.9 Teacher 3, Science, second unit, class 1

1	TCH:	And here in the last picture you see like some people, what are they doing in the last picture?
2	STU:	They are..
3	TCH:	What are they doing, Gloria?
4	STU:	<L1 Arrancar L1>, they are <L1 arrancar L1> <x…x>
5	TCH:	In this one. What do you think they are doing?
6	STU:	Ah! <L1 ¿En ésa? L1>. they are doing their house

The great majority of didactic or *re-routing recasts* focus on *language* in the case of all teachers (all percentages over 90%) and scarcely solely on content (0% for Teachers 2 and 4, 0.9% for Teacher 1, 1.2% for Teacher 3). **Extract 8.10** shows a *didactic recast* focusing on language (specifically, on feminine/masculine pronouns), which the student takes up in his next turn.

Extract 8.10 Teacher 1, Science, second unit, class 5

1	TCH:	What did she write to remember in her house?
2	STU:	He wrote..
3	TCH:	She wrote
4	STU:	She wrote eh.. these are the correct

Regarding the type of *prompt* there is more variety across teachers: Teacher 1 focuses on *content* in 43.3% of the cases; Teacher 2 mainly focuses on both *content + language* (40.8%); Teacher 3 clearly focuses more on *content* (66.2%); and, finally, in Teacher 4's prompts, the main focus is also *content* (48.5%).

	AfL		Non-AfL	
	Teacher 1	Teacher 2	Teacher 3	Teacher 4
EVALUATION				
Positive evaluation	**87.3%**	**95.0%**	**89.8%**	**86.2%**
With recast	23.8%	35.6%	29.0%	21.4%
Content	*2.9%*	*4.1%*	*3.5%*	*5.7%*
Language	*92.9%*	*93.2%*	*93.0%*	*89.7%*
Content + Language	*4.1%*	*2.7%*	*3.5%*	*4.6%*
With no recast	76.2%	64.4%	71.0%	78.6%
Negative evaluation	**12.7%**	**5.0%**	**10.2%**	**13.8%**
With explicit correction	66.4%	63.6%	59.1%	53.8%
Content	*29.6%*	*42.9%*	*76.9%*	*48.6%*
Language	*56.1%*	*14.3%*	*15.4%*	*40.0%*
Content + Language	*14.3%*	*42.9%*	*7.7%*	*11.4%*
With no explicit correction	33.6%	36.4%	40.9%	46.2%
RE-ROUTE				
Prompt	78.9%	59.0%	76.4%	80.2%
Content	*43.3%*	*36.7%*	*62.2%*	*48.5%*
Language	*32.2%*	*22.4%*	*16.5%*	*20.9%*
Content + Language	*24.4%*	*40.8%*	*21.2%*	*30.6%*
Recast	21.1%	41.0%	23.6%	19.8%
Content	*0.9%*	—	*1.2%*	—
Language	*94.3%*	*91.2%*	*96.5%*	*92.2%*
Content + Language	*4.7%*	*8.8%*	*2.3%*	*7.8%*

Table 8.3 Evaluation and re-route types.

Summing up, the most common type of feedback move is *evaluation*, followed by *expansion*, *re-route*, *meta-feedback*, and finally, *revision*. This is the case for all teachers, although we will see later if differences between them are significant. Regarding the type of *evaluation*, *positive evaluation with no recast* is the most common for all teachers. When there is *negative evaluation*, Teachers 1 and 2 tend to offer *explicit correction* more frequently than Teachers 3 and 4. When the feedback move is *re-route*, *prompts focusing on content* are the most common ones. When *re-routing* is done through *recasts*, most of them *focus on language* in the case of all teachers.

Moving on to the comparison between AfL and Non-AfL teachers, **Table 8.4** below illustrates the differences in their use of feedback moves. The results show that *evaluation* is the most frequent type for both AfL and Non-AfL teachers. As for the rest of main feedback types, *expansion* and *meta-feedback* appear more frequently in AfL classes, whereas *re-route* and *revision* are more frequent in Non-AfL classes. Within the different subtypes of *evaluation*, significant differences only appear with *negative evaluation* and *explicit correction*. Within the former, AfL teachers make more *negative evaluations including explicit corrections*. Within the latter, Non-AfL teachers' *explicit corrections* focus more on *content*, as opposed to AfL teachers, focusing more on *language*. In the same way, *prompts* are aimed more frequently at *content* in Non-AfL classes, whereas in AfL classes they are more directed towards *language*.

		AfL		Non-AfL		
	%	N	%	N	*T*	χ^2
Evaluation	42.2%	1393	41.6%	1148	0.53	0.28
Expansion	26.6%	876	23.2%	641	3.00***	9.01***
Revision	1.3%	42	6.1%	169	10.34***	105.01***
Re-route	17.8%	586	22.5%	621	4.58***	20.95***
Meta-feedback	12.2%	401	6.6%	183	7.30***	52.85***
Total		3298		2762		
Evaluation						
Positive	88.5%	1229	88.3%	1012	0.14	0.02
Negative	11.5%	160	11.7%	134	0.14	0.02
Total		1389		1146		
Positive evaluation						
With recast	25.8%	314	25.9%	259	0.04	0.00
With no recast	74.2%	902	74.1%	741	0.04	0.00
Total		1216		1000		
With recast						
Content	3.2%	10	4.2%	11	0.67	0.45
Language	93.0%	292	91.9%	238	0.50	0.25
Content + Language	3.8%	12	3.9%	10	0.02	0.00
Total		314		259		
Negative evaluation						
With explicit correction	66.2%	106	56.5%	74	1.71*	2.91*
With no explicit correction	33.8%	54	43.5%	57	1.71*	2.91*
Total		160		131		
With explicit correction						
Content	30.5%	32	63.5%	47	4.61***	19.22***
Language	53.3%	56	27.0%	20	3.61***	12.30***
Content + Language	16.2%	17	9.5%	7	1.30	1.69
Total		105		74		
Re-route						
Prompt	76.1%	446	77.9%	484	0.76	0.57
Recast	23.9%	140	22.1%	137	0.76	0.57
Total		586		621		
Prompt						
Content	42.6%	190	56.4%	273	4.24***	17.69***
Language	31.2%	139	18.4%	89	4.57***	20.48***
Content + Language	26.2%	117	25.2%	122	0.36	0.13
Total		446		484		
Recast						
Content	0.7%	1	0.7%	1	0.02	0.00
Language	93.6%	131	94.9%	130	0.47	0.22
Content + Language	5.7%	8	4.4%	6	0.51	0.26
Total		140		137		
Note: *$p < 0.01$; ***$p < 0.001$						

Table 8.4 Feedback types: comparison between AfL and Non-AfL teachers.

Extract 8.11 below comes from an AfL Science lesson. Paying attention only to feedback types, it seems clear that the kinds of feedback used by this AfL teacher are quite varied: there is *evaluation* (turns 4, 13, 27), *re-routing* (using mainly *prompts*, in turns 11, 13, 25), *meta-feedback* (turn 6), and *expansion* (turns 14, 20, 23).

Extract 8.11 Teacher 1, Science, second unit, class 2

1	TCH:	Your level is very high and here you are just changing the same idea and write it in a different way, ok?
2		Eh.. another sentence, Pilar
3	STU1:	Eh.. bones can fracture
4	TCH:	Bones can fracture
5	STU1:	And
6	TCH:	One minute, if you say bones can fracture and then you can write “break” here ((writing the sentence and writing “break” under “fracture”)), that is good because maybe somebody doesn't remember this word for the examination and they will put “break”. So if you put both words, it's better to study, and then on the examination you could put the most difficult words that you want.
7		Bones can fracture
8	STU1:	And muscles can tear ((pronouncing the word incorrectly))
9	TCH:	And muscles can too?
10	STU:	Tear ((pronouncing wrongly))
11	TCH:	Tear ((correcting pronunciation)) Ah.. shall I put “and”.. or “but”?
12	STU1	But
13	TCH:	But. So before “but” there's always a comma, if you don't know that I'm circling it..
14		But muscles.. Muscles, muscles ((writing the second part of the sentence)) Can you tell me a verb that rhymes with “tear”? ((No answer for some seconds)).
15		Muscles tear. Eh.. we started the sense of hearing and I can hear you.
16		Does this verb and this verb go together? ((writes “hear” on the board))
17	STU:	Yes
18	STU:	No
19	TCH:	((Says “No” with her head)) But they are spelt the same
20	STU:	Yes
21	TCH:	So what's the problem?
22	STU:	The “h” and the “t”
23	TCH:	They're spelt the same but they sound different, so this is not a good example to remember the spelling. What word do you use for your clothes?
24	STU:	tear
25	TCH:	Everyday I..
26	STU:	((Some)) Wear
27	TCH:	Wear.

On the contrary, **Extract 8.12** below, coming from a Non-AfL class, shows how the feedback types used by the teacher are not so varied. They include mainly *evaluation* (turns 7, 9, 11, 13, 16, 18, 21), and some *expansion* (turn 22) and *re-route* (turn 24). *Evaluation* may be in connection with the high frequency of questions for facts, while *re-route* may be partly related to the use of questions for reasons, as will be explained later in this chapter.

Extract 8.12 Teacher 3, Science, first unit, class 2

1	TCH:	Ok, so.. what are the parts of a bird's body? Natalia
2	STU1:	the trunk, the tail
3	TCH:	Natalia, you need to speak up a little bit because I don't hear anything with the construction outside
4	STU1:	the trunk, the tail
5	TCH:	So we said.. Sorry, Natalia
6	STU1:	the trunk
7	TCH:	So they have a trunk
8	STU1:	a tail
9	TCH:	Tail
10	STU1:	and limbs
11	TCH:	Limbs and?
12	STU:	A head!
13	TCH:	And a head.
14	TCH:	What's the name of the front limbs? The front limbs are called? Jessica
15	STU2:	Wings
16	TCH:	Wings, and the back? What's the name of the back limbs, Alejandro?
17	STU3:	Legs
18	TCH:	Legs, ok? Very good.
19		They body is covered, the body of a bird, tell me what is it covered.. by? Mauro
20	STU4:	Feathers
21	TCH:	By feathers.
22		And why are feathers important for a bird? There are several reasons, ok? Why are the feathers important... to a bird? Let's see... eh Yolanda
23	STU5:	for use the wings
24	TCH:	Ok, so feathers.. the body of a bird is covered by feathers and they have feathers in their wings, so that means that feathers.. what's an important job they have? María Belén
25	STU6:	can fly

This section has shown, both quantitatively and qualitatively, how AfL teachers and Non-AfL teachers use feedback types in significantly different ways. As a summary, AfL teachers use more *expansion* and *meta-feedback*, thus placing the process of learning, assessment, and building on students' ideas and contributions at the fore. Likewise, AfL teachers also tend to focus more on *language* when they use *explicit corrections* and *prompts*. Non-AfL teachers, on their part, use more *re-route* and *revision*, and their *prompts* and *explicit corrections* focus more on *content*. As explained in Chapter 2, most CLIL curricula are content-based (Dalton-Puffer and Smit 2007), and therefore content is supposed to have more weight than language. In relation to individual differences, *evaluation* is very frequent in all teachers, the same as *revision* is the least used type of feedback. Teacher 2 is the one who most uses *expansion* and Teacher 1 the one who most frequently offers *meta-feedback*.

8.3 TEACHER FEEDBACK AND QUESTION TYPE

Another interesting feature to look into is the relationship between feedback type and question type in order to discover whether certain types of questions are more likely to be followed by certain types of feedback. As part of the same IRF pattern, the question type could influence the feedback type to some extent, in the same way it can also influence students' responses (see Chapter 7). **Table 8.5** presents each type of question and the percentages of the main feedback types following them. The results show that *evaluation* is very frequent in general, regardless of the question type, although it is higher when a *question for fact* (44.1%), *for opinion* (46.9%), or *language question* (44.1%) has been previously asked. *Expansion* is the second type of feedback more frequently used, and especially so when the teacher has previously asked a *meta-cognitive question* (35.2%) or a *question for reason* (39.1%). In fact, *questions for reasons* are mainly followed by *expansion* feedback moves. *Re-route* is also a frequent type of feedback, used mainly with *questions for facts*, *for explanations*, and *language questions* (20.1%, 24.1%, 20.3%, respectively). *Meta-feedback*, less frequent than *evaluation*, *expansion*, and *re-route*, is mainly preceded by *meta-questions* (41.1%). Finally, *revision* is infrequent in general and it is mainly found with *questions for facts* (5%).

	Fact	Explanation	Reason	Opinion	Meta-cognitive	Meta-question	Language question
Evaluation	44.1%	38.0%	35.8%	46.9%	38.5%	31.6%	44.1%
Expansion	24.1%	29.2%	39.1%	29.5%	35.2%	11.7%	26.1%
Revision	5.0%	1.7%	1.0%	0.7%	0.8%	0.8%	3.1%
Re-route	20.1%	24.1%	16.9%	17.0%	18.4%	14.8%	20.3%
Meta-feedback	6.6%	7.0%	7.3%	5.9%	7.0%	41.1%	6.4%
Total	3327	644	302	288	244	494	719

Table 8.5 Feedback type depending on type of question.

Next, some examples instantiating question type together with feedback type will be offered. Each type of question will be exemplified as to what type of feedback most frequently follows. **Extract 8.13** illustrates the frequent occurrence of *questions for facts* followed by *evaluation* in the feedback move. The teacher asks questions for facts related to where democracy was born and where Athens is and she evaluates students' responses positively.

Extract 8.13 Teacher 4, Citizenship, first unit, class 1

1 TCH: Ok, so the first democracy was born in….? (question for fact)
2 STU: Athens
3 TCH: Athens. (positive evaluation) Athens if the capital city of? (question for fact)
4 STU: Greece!!
5 STU: Greece!
6 TCH: Greece, fine (positive evaluation)

After *evaluation*, the next frequent feedback types which follow questions for *facts are expansion* (illustrated in **Extract 8.14**) and *re-route* (illustrated in **Extract 8.15**). In **Extract 8.14**, the feedback move *expansion* is, at the same time, another initiation (*question for reason*). In the case of **Extract 8.15**, the feedback move *re-route* is specifically a *prompt focusing on content*.

Extract 8.14 Teacher 3, second unit, class 1

1	TCH:	Do you know by chance which are those periods Prehistory is divided into? Nadia (question for fact)
2	STU1:	the Stone Age and the Metal Age
3	TCH:	Very good. The Stone Age and the Metal Ages, ok? Do you know what's the difference.. why are they called the Stone Age and the Metal Ages, Mario? (expansion)
4	STU2:	eh.. the when- the Stone Age is of.. of.. <L1 piedra L1> and

Extract 8.15 Teacher 3, Science, first unit, class 5

1	TCH:	My question is what are short conversations? (question for fact)
2	STU:	we don't have time
3	TCH:	We don't have time? My question is what are short conversations? Short conversations? (re-route: prompt)
4	STU:	Eh… a conversation that… is very short.

Questions for explanations are mainly followed by *evaluation*, *expansion*, and *re-route* (in fact, the highest percentage of *re-route* is after *questions for explanations*). In **Extract 8.16**, the teacher asks the learners to explain what another classmate was trying to say and, after the explanation is offered, the teacher evaluates the student's contribution positively.

Extract 8.16 Teacher 2, Citizenship, second unit, classes 1 and 2

1	TCH:	Do you know how to explain what Guillermo is saying?(question for explanation)
2	STU:	Yes! That eh, do a sport is very healthy for your body
3	TCH:	Very healthy for your body, good, (evaluation)

In 126, there are two *questions for explanations*, the second one being *expansion* at the same time. In turn 5 there is another teacher *expansion*.

Extract 8.17 Teacher 1, Science, second unit, class 5

1	TCH:	The muscle, ok, when you contract the biceps what happens? (question for explanation)
2	STU:	Em, that your biceps…
3	TCH:	That your biceps contract which means %x….x% can you explain this? (question for explanation + expansion) ((Flexes))
4	STU:	When you get up, when you…when your…arm goes up
5	TCH:	Ahh! See? You were looking and looking and you found it (positive evaluation). So continue (expansion). ((goes to another group))
6	STU:	When your arm goes up, your muscles, your muscle contracts.

In **Extract 8.18**, after the teacher asks the *question for explanation* "how do you know that?", she has to use *re-route* to arrive at the answer (first, with a *recast focusing on language*; secondly, by repeating the same question; and thirdly, with a *prompt*).

Extract 8.18 Teacher 1, Citizenship, second unit, class 1

1	TCH:	She wants to hug the boy, and how do you know that? What is the tree doing with its branches? Ricardo (question for explanation)
2	STU1:	He is going to…
3	TCH:	She is… (re-route: recast)
4	STU1:	She is going to…to hug the boy.
5	TCH:	But how do you know that? (question for explanation)
6	STU2:	Because the
7	TCH:	The br…(re-route: prompt)
8	STU3:	The branches is like go…is like this ((opens arms up))

When *questions for reasons* are asked, they are normally followed by *expansion* in the feedback move, as illustrated in **Extract 8.19**. In this case, the *expansion* is, at the same time, a *meta-cognitive question* and it is preceded by a short *positive evaluation*.

Extract 8.19 Teacher 2, Citizenship, second unit, classes 1 and 2

1	TCH:	About being healthy, good. Um, why are we looking at being healthy? Matías? (question for reason)
2	STU1:	To have a good diet and…to be good.
3	TCH:	((correction)) To be good, to feel good. Good, why else do you think, Verónica? (expansion)
4	STU2:	to know how eh, we are eh, how is the level of our health.
5	TCH:	Good, to know about ourselves and our own health, good, that's a good suggestion

Surprisingly, *questions for opinions* are also often followed by *evaluations* (see **Extract 8.20**). One possible *explanation* is that, sometimes, the *question for opinion* actually represents a *question for fact*, in which the teacher is expecting a certain answer.

Extract 8.20 Teacher 3, Science, second unit, class 1

1	TCH:	Do you think this is the horse? (question for opinion)
2	STU:	yes
3	TCH:	Yes, I would think so as well. That's a horse (evaluation)

Extract 8.21 is an illustration of the frequent sequence *question for opinion* and *expansion* in the feedback move.

Extract 8.21 Teacher 4, Citizenship, first unit, class 4

1	TCH:	Do you think wearing uniform is a positive or a negative thing, Adrián? (question for opinion)
2	STU:	It's a negative thing
3	TCH:	Ok, come on (expansion)
4	STU:	That, em, in the schools, em, when you you see that, that they wear a uniform, some people only go to that schools because the uniforms are cute but

Most *meta-questions* are followed by *meta-feedback*. It is expected that if the teacher's question has to do with getting students to reflect on their learning and how to improve, the feedback move in the same exchange is related to the same processes (see **Extract 8.22**).

Extract 8.22 Teacher 1, Citizenship, first unit, class 3

1	TCH:	is it perfect? (meta-question)
2	STU:	no
3	STU:	Comma!
4	TCH:	Comma. Almost perfect, almost perfect but this person can't have a <L1 bien L1>, because this person continues to have some little problems.(meta-feedback)

Meta-cognitive questions are almost equally followed by *evaluation* and *expansion*. **Extract 8.23** illustrates both cases, since the feedback move is composed of *positive evaluation* and *expansion*.

Extract 8.23 Teacher 2, Citizenship, second unit, classes 1 and 2

1	TCH:	Yes, ok, Alfonso?
2	STU1:	And they talk about don't eat a lot of meat
3	TCH:	Not eating a lot of meat, ok, why do you think that we should not eat a lot of meat? (meta-cognitive question)
4	STU1:	Eh, because, it's bad for the heart ((pronounced incorrectly))
5	TCH:	Bad for your health, yeah, (positive evaluation) who can expand on that? Who can tell me a bit more about that? Alejo?(expansion)
6	STU2:	For to not have %L1…grasa…L1%

Language questions are most of the times followed by *evaluation* and *re-route*. This is expected given that teachers often looked for specific terms or the self-correction of specific grammar aspects when asking *language questions*. **Extracts 8.24 and 8.25** include examples of both types of feedback (*evaluation* and *re-route*, respectively, following *language questions*). *Language questions* followed by *re-route* when the response has been incorrect are very frequent in the data (see **Extract 8.25**, with two *prompts focusing on language*).

Extract 8.24 Teacher 3, Science, second unit, class 1

1	TCH:	What is pottery? (language question)
2	STU:	<L 1Cerámica L1>
3	TCH:	<L1 Cerámica L1>. (evaluation) The wheel? (language question)
4	STU:	<L1 La rueda L1>
5	TCH:	<L1 La rueda L1>. (evaluation)

Extract 8.25 Teacher 1, Science, second unit, class 1

1	TCH:	Can you tell me another word for join? (language question)
2	STU:	Joins
3	TCH:	Join. (re-route)
4	STU:	Joint
5	TCH:	I need a verb. The place where two bo- bo- bo- bones (re-route)
6	STU:	Connect
7	STU:	meet
8	TCH:	Meet, connect.

In order to be able to run comparisons of feedback types in relation to question types in the clearest possible way, I have divided question types into four groups, depending on the type of feedback that follows, so that comparisons can be reliable but more feasible at the same time. One group includes *questions for facts* and *language questions*, another one *questions for explanations* and *for opinions*, the third one *questions for reasons* and *meta-cognitive questions* and, finally, *meta-questions* alone in the fourth group. The groups have been decided according to the similarities in the types of feedback that follow them (see **Table 8.5** above).

The following tables will compare the four groups of questions explained above with regard to the feedback types that follow them in two-way comparisons. **Table 8.6** below presents the comparison regarding the type of feedback following *questions for facts/language questions* versus *questions for explanations/opinions*. Results indicate that, when a *question for explanation/opinion* is asked, *expansion* (29.3%), *positive evaluation* (92.4%), *positive evaluation with recasts* (31.5%), and *didactic recasts* (34.8%) are significantly more frequent. On the other hand, when a *question for fact/language question* is asked, *revision* (4.6%),

negative evaluation (12.5%), *positive evaluation without recasts* (77.2%), and *prompts* (81.7%) are all significantly more frequent.

	Facts / Language		Opinions / Explanations			
	%	N	%	N	*T*	χ^2
Evaluation	44.2%	1784	40.8%	380	1.89*	3.55
Expansion	24.4%	987	29.3%	273	3.07***	9.43
Revision	4.6%	186	1.4%	13	4.51***	20.31
Re-route	20.2%	815	21.9%	204	1.17	1.36
Meta-feedback	6.6%	267	6.7%	62	0.05	0.00
Total		4039		932		
Evaluation						
Positive	87.5%	1556	92.4%	351	2.71***	7.30***
Negative	12.5%	223	7.6%	29	2.71***	7.30***
Total		1779		380		
Positive evaluation						
With recast	22.8%	351	31.5%	110	3.43***	11.70***
With no recast	77.2%	1188	68.5%	239	3.43***	11.70***
Total		1539		349		
Re-route						
Prompt	81.7%	666	65.2%	133	5.19***	26.31***
Recast	18.3%	149	34.8%	71	5.19***	26.31***
Total		815		204		
Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$						

Table 8.6 Feedback types following questions for facts/language vs. questions for explanations/opinions.

Table 8.7 compares feedback types following *questions for facts/language questions* with those following *questions for reasons/meta-cognitive questions*.

	Facts / Language		Reasons / Meta-cognitive			
	%	N	%	N	*T*	χ^2
Evaluation	44.2%	1784	37.0%	202	3.18***	10.08***
Expansion	24.4%	987	37.4%	204	6.49***	41.79***
Revision	4.6%	186	0.9%	5	4.06***	16.40***
Re-route	20.2%	815	17.6%	96	1.43	2.04
Meta-feedback	6.6%	267	7.1%	39	0.47	0.22
Total		4039		546		
Evaluation						
Positive	87.5%	1556	92.1%	186	1.91*	3.64*
Negative	12.5%	223	7.9%	16	1.91*	3.64*
Total		1779		202		
Positive evaluation						
With recast	22.8%	351	35.3%	65	3.76***	14.06***
With no recast	77.2%	1188	64.7%	119	3.76***	14.06***
Total		1539		184		
Re-route						
Prompt	81.7%	666	70.8%	68	2.56***	6.50***
Recast	18.3%	149	29.2%	28	2.56***	6.50***
Total		815		96		
Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$						

Table 8.7 Feedback types following questions for facts/language vs. questions for reasons/meta-cognitive questions.

In **Table 8.7** we can see significant differences in feedback types when we compare ques*tions for facts/language questions* and *questions for reasons/meta-cognitive questions*. *Evaluation* (44.2%), *revision* (4.6%) and *re-route* (20.2%) are significantly more frequent after the former; *expansion* (37.4%) and *meta-feedback* (7.1%) after the latter. In addition, *positive evaluation* (92.1%), *positive evaluation with recast* (35.3%), and *didactic recasts* (29.2%) are significantly more frequent with *questions for reasons/meta-cognitive questions*; *negative evaluation* (12.5%), *positive evaluation without recast* (77.2%), and *prompts* (81.7%) are significantly more frequent after *questions for facts/language questions*.

When feedback types following *questions for facts/language questions* are compared with feedback types following *meta-questions*, again, a lot of significant differences appear (see **Table 8.8**): *evaluation* (44.2%), *expansion* (24.4%), *revision* (4.6%), and *re-route* (20.2%) are all more frequent with *questions for facts/language questions*, as opposed to *meta-feedback*, which is more present with *meta-questions* (41.1%). *Positive evaluation with recast* (33.3%), *didactic recasts* (35.6%), and *prompts focusing on language* (46.8%) are significantly more frequent with *meta-questions*, whereas *positive evaluation with no recast* (77.2%), *prompts* (81.7%), and *prompts focusing on content + language* (27.6%) are more frequently used with *questions for facts/language questions*.

	Facts / Language		Meta-questions			
	%	N	%	N	*T*	χ^2
Evaluation	44.2%	1784	31.6%	156	5.35***	28.50***
Expansion	24.4%	987	11.7%	58	6.35***	40.00***
Revision	4.6%	186	0.8%	4	3.98***	15.79***
Re-route	20.2%	815	14.8%	73	2.86***	8.15***
Meta-feedback	6.6%	267	41.1%	203	25.35***	563.17***
Total		4039		494		
Positive evaluation						
With recast	22.8%	351	33.3%	42	2.68***	7.16***
With no recast	77.2%	1188	66.7%	84	2.68***	7.16***
Total		1539		126		
Re-route						
Prompt	81.7%	666	64.4%	47	3.59***	12.72***
Recast	18.3%	149	35.6%	26	3.59***	12.72***
Total		815		73		
Prompt						
Content	48.6%	324	40.4%	19	1.09	1.19
Language	23.7%	158	46.8%	22	3.55***	12.40***
Content + Language	27.6%	184	12.8%	6	2.23**	4.96**
Total		666		47		

Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$

Table 8.8 Feedback types following questions for facts/language questions vs. meta-questions.

Table 8.9 below focuses on the comparison between feedback types following *questions for explanations/opinions* and feedback types following *questions for reasons/meta-cognitive questions*.

	Explanations / Opinions		Reasons / Meta-cognitive			
	%	N	%	N	*T*	χ^2
Evaluation	37.0%	202	40.8%	380	1.43	2.06
Expansion	37.4%	204	29.3%	273	3.21***	10.26***
Revision	0.9%	5	1.4%	13	0.81	0.66
Re-route	17.6%	96	21.9%	204	1.99**	3.95**
Meta-feedback	7.1%	39	6.7%	62	0.36	0.13
Total		546		932		

Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$

Table 8.9 Feedback types following questions for explanations/opinions vs. questions for reasons/meta-cognitive questions.

Few significant differences appear when questions for *explanations/opinions* and *questions for reasons/meta-cognitive questions* are compared. The only difference is that *expansion* is more used with the latter and *re-route* with the former.

Table 8.10 below shows the comparison between *questions for explanations/opinions* and *meta-questions*. As it has been pointed out before, *meta-questions* are different from the rest in relation to the feedback type that may follow them, and so there are many significant differences. All feedback types are significantly more abundant when *questions for explanations or for opinions* are asked, except for *meta-feedback* (41.1%), which is the type of feedback that a *meta-question* normally gets. Other significant differences are that *positive evaluation* (92.4%) is more frequent when there is a *question for explanation or for opinion*, and *negative evaluation* (16.7%) and *prompts focusing on language* (46.8%) are more numerous with *meta-questions*.

	Explanations/Opinions		Meta-questions			
	%	N	%	N	*T*	χ^2
Evaluation	40.8%	380	31.6%	156	3.42***	11.63***
Expansion	29.3%	273	11.7%	58	7.62***	55.80***
Revision	1.4%	13	0.8%	4	0.97	0.94
Re-route	21.9%	204	14.8%	73	3.24***	10.43***
Meta-feedback	6.7%	62	41.1%	203	17.53***	253.12***
Total		932		494		
Evaluation						
Positive	92.4%	351	83.3%	130	3.15***	9.80***
Negative	7.6%	29	16.7%	26	3.15***	9.80***
Total		380		156		
Prompt						
Content	52.6%	70	40.4%	19	1.44	2.07
Language	24.1%	32	46.8%	22	2.98***	8.56***
Content + Language	23.3%	31	12.8%	6	1.54	2.36
Total		133		47		
Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$						

Table 8.10 Feedback types following questions for explanations/opinions vs. meta-questions.

Significant differences are also found when feedback types following *questions for reasons/meta-cognitive questions* are compared with feedback types following *meta-questions* (**Table 8.11**): *evaluation* (37%), *expansion* (37.4%), and *positive evaluation* (92.1%) are more frequent with the former, *meta-feedback* (41.1%), *negative evaluation* (16.7%), and *prompts focusing on language* (46.8%) are significantly more abundant with the latter.

	Reasons / Meta-cognitive		Meta-questions			
	%	N	%	N	*T*	χ^2
Evaluation	37.0%	202	31.6%	156	1.84*	3.37*
Expansion	37.4%	204	11.7%	58	9.94***	90.34***
Revision	0.9%	5	0.8%	4	0.18	0.03
Re-route	17.6%	96	14.8%	73	1.22	1.50
Meta-feedback	7.1%	39	41.1%	203	14.11***	167.43***
Total		546		494		
Evaluation						
Positive	92.1%	186	83.3%	130	2.57`***	6.50***
Negative	7.9%	16	16.7%	26	2.57***	6.50***
Total		202		156		
Prompt						
Content	52.9%	36	40.4%	19	1.32	1.74
Language	23.5%	16	46.8%	22	2.67***	6.81***
Content + Language	23.5%	16	12.8%	6	1.44	2.08
Total		68		47		
Note: *$p < 0.01$; ***$p < 0.001$						

Table 8.11 Feedback types following questions for reasons/meta-cognitive questions vs. meta-questions.

As shown above both with the quantitative analysis and the examples used to illustrate the main findings, question types determine, to some extent, the type of feedback offered in the same exchange. For instance, it is very common for *questions for facts* to be followed by *evaluation*. In contrast, *questions for explanations* are followed by a variety of feedback types, such as *evaluation*, *expansion*, and *re-route*. In the case of *questions for reasons*, *expansion* is the most frequent feedback type used after them. *Questions for opinions* are mainly followed by *evaluation* and *expansion*. As for *meta-cognitive questions*, *evaluation* and *expansion* are the two feedback types following this type of question more frequently. Regarding *meta-questions*, they are usually followed by *meta-feedback*. Finally, *evaluation* and *expansion* are the most frequent types of feedback following *language questions*. Two-way comparisons were made to prove if the differences observed in **Table 8.5** were significant, and they turned to be so indeed. In the next chapter, the implications of these significant differences will be addressed.

8.4 TEACHER FEEDBACK AND STUDENT RESPONSE

After having focused on types of feedback, comparing AfL and Non-AfL teachers, and the relationship between question type and feedback type, the present section addresses the link between types of responses and types of feedback. We consider that feedback is both related to the type of question and to the with type of response. **Table 8.12** below presents the percentages of students' responses and feedback types. From all students' responses consisting in a *minimal response*, half of them get *evaluation* (49.2%), 21.4% are followed by *expansion*, 14.8% by *re-route*, 2.1% by *revision*, and 12.5% by *meta-feedback*. It is very interesting that *minimal response* is the type which is mostly followed by *expansion* and *meta-feedback* (see **Extracts 8.26 and 8.27**, respectively).

Extract 8.26 Teacher 2, Citizenship, second unit, classes 1 and 2

1	TCH:	You know, when I first saw it, I thought he was eating ice cream. Say if it was rice, would that be healthy or not?
2	STU:	((All)) Yes
3	TCH:	And if it was ice cream?
4	STU:	((All)) No
5	TCH:	Why is it not so healthy to eat ice cream? Bárbara?
6	STU1:	Because it's like a bomb to your stomach. Because it's so, em, cold
7	TCH:	It's so cold
8	STU1:	And, no it's very good to eat

Extract 8.27 Teacher 1, Science second unit, class 1

1	STU:	they said that I'm improving but.. but that I have to do more effort working at my level
2	TCH:	Working at my level. Do you know your level?
3	STU:	yes
4	TCH:	Ok? Because some people say "how do I work at my level? I don't know my level". I think more or less, everybody knows their level.

It seems very reasonable for *expansion* to follow *minimal responses*. After a *yes/no* answer, the teacher may want to get the student to articulate a longer response and relies on either *expansion* or *meta-feedback* to do so.

Truncated responses are followed, in their majority, by *re-route* (especially *prompts*), then *expansion*, *evaluation*, *meta-feedback*, and *revision*. It is also expected that *truncated responses* are the ones that are followed by *re-route* more frequently (as illustrated in **Extract 8.28**). When students cannot respond to the question appropriately, the teacher scaffolds them into an answer through the use of re-routing until the correct answer is given, or as much as they can before ending up asking another student or providing the answer themselves.

Extract 8.28 Teacher 3, Science, first unit, class 4

1	TCH:	Where do they live?
2	STU1:	in..
3	TCH:	Baby frogs, do they live on land?
4	STU:	In the rivers

	Minimal	Truncated	T-Unit			
			Total	*One-phrase*	*One-clause*	*More than one-clause*
Evaluation	49.2%	13.4%	65.6%	72.1%	57.8%	60.1%
Expansion	21.4%	15.2%	9.0%	7.3%	11.8%	8.8%
Revision	2.1%	2.7%	1.3%	1.4%	1.7%	—
Re-route	14.8%	63.4%	19.6%	15.6%	24.1%	24.2%
Prompt	*98.2%*	*87.3%*	*68.6%*	78.9%	62.4%	57.1%
Recast	*1.8%*	*12.7%*	*31.4%*	21.1%	37.6%	42.9%
Meta-feedback	12.5%	5.4%	4.4%	3.6%	4.7%	6.9%

Table 8.12 Feedback types according to student responses.

Out of all the *T-unit responses*, more than half (65.6%) are followed by *evaluation*, 9% by *expansion*, 1.3% by *revision*, 19.6% by *re-route*, and 4.4% by *meta-feedback*. Within *T-unit responses* types, *one-phrase* are mainly followed by *evaluations* (72.1%) and then *re-route* (15.6%). In **Extract 8.29**, after the teacher's *language question*, the student answers with *one phrase* ("weird"). Subsequently, the teacher's feedback consists in *positive evaluation* (turn 3).

Extract 8.29 Teacher 4, Arts, first unit, first class

1 TCH: what's another, sorry, what's another %x…x% we know for strange?
2 STU: Weird
3 TCH: Weird, for example, ok?

For *one-clause responses*, *evaluation* is again the most common feedback type (57.8%), although the percentage is lower than for *one-phrase responses*. *Re-route* (24.1%) and *expansion* (11.8%) also follow *one-clause responses* quite often. The same applies to *more-than-one-clause responses*, which are mainly followed by *evaluation* (60.1%) and then by *re-route* (24.2%). *Meta-feedback* is not very commonly used after *more-than-one-clause responses*, and *revision* is not found at all. *Revision* is not used after students' responses very often but, rather, after another type of teacher feedback in the same turn, as instantiated in **Extract 8.30**. In this extract, the teacher is reviewing the characteristics of pop art, which they studied in the previous lesson. In turn 3, after the *positive evaluation* "bright colours, ok?", she goes on reviewing the content already explained.

Extract 8.30 Teacher 3, Arts, second unit, class 2

1 TCH: What type of colours did they use..?
2 STU: bright colours
3 TCH: Bright colours, ok? And the lines, very clear lines, ok? And it was very easy to understand and to recognise, because they weren't looking for second meanings, ok?

In this section, teacher feedback types and the preceding student response types have been analysed. The most interesting results have shown how, after *minimal responses*, the most frequent feedback types are *expansion* and *meta-feedback*. Likewise, after *truncated responses*, it is very common to find *re-route*, in order to help the student find the most appropriate answer.

8.5 TEACHER FEEDBACK AND SUBJECT TYPE

This section focuses on the analysis of feedback types across subjects. The different types of feedback used by the same teachers in two different subjects will also be examined, in this way exploring whether there are differences in the way teachers use feedback in different subjects.

Table 8.13 below shows the comparison between Science and Citizenship. In Science, *evaluation* is the most frequent type of feedback (41.4%), followed by *expansion* (27%), *re-route* (20.2%), *meta-feedback* (5.9%), and *revision* (5.5%). In Citizenship, the main feedback types are also *evaluation* (44.4%), *expansion* (24.2), *re-route* (20.2%), *meta-feedback* (10.4%),

and *revision* (0.8%). When these subjects are compared, significant differences appear in all the categories except for *re-route*: *evaluation* and *meta-feedback* are more frequent in Citizenship, whereas *expansion* and *revision* are more abundant in Science. As for subtypes of feedback, *negative evaluation with explicit correction*, *re-routing recasts*, *prompts focusing on both content + language* and *re-routing recasts focusing on both content + language* are more used in Citizenship. *Prompts* in general and, in particular, *prompts focusing on language* are more repeated in Science lessons.

	Science		Citizenship			
	%	N	%	N	*T*	χ^2
Evaluation	41.4%	1254	44.4%	1049	2.21**	4.87**
Expansion	27.0%	817	24.2%	571	2.34**	5.47***
Revision	5.5%	166	0.8%	20	9.32***	85.58***
Re-route	20.2%	613	20.2%	477	0.05	0.00
Meta-feedback	5.9%	178	10.4%	245	6.11***	37.06***
Total		3028		2362		
Negative evaluation						
With explicit correction	61.2%	85	71.6%	78	1.72*	2.94*
With no explicit correction	38.8%	54	28.4%	31	1.72*	2.94*
Total		139		109		
Re-route						
Prompt	79.0%	484	73.4%	350	2.16**	4.65**
Recast	21.0%	129	26.6%	127	2.16**	4.65**
Total		613		477		
Prompt						
Content	50.0%	242	47.4%	166	0.73	0.54
Language	27.5%	133	21.1%	74	2.09**	4.37**
Content + Language	22.5%	109	31.4%	110	2.90***	8.32***
Total		484		350		
Recast						
Content	1.6%	2	—	—	0.00	1.98
Language	96.9%	125	92.9%	118	1.45	2.11
Content + Language	1.6%	2	7.1%	9	2.20**	4.77**
Total		129		127		
Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$						

Table 8.13 Comparison feedback types Science vs. Citizenship.

Extracts 8.31 and 8.32 show two types of frequent feedback types in Science (**Extract 8.31**) and Citizenship (**Extract 8.32**), other than *evaluation*. Specifically, **Extract 8.31** illustrates *expansion* in a Science lesson (turn 3) and **Extract 8.32** instantiates *meta-feedback* in a Citizenship lesson (turn 3).

Extract 8.31 Teacher 1, Science, second unit, class 3

1 TCH: What about you?
2 STU: I completed, I explain to my mother but my mother doesn't have a comment
3 TCH: Why not?
4 STU: eh because she don't have time and she sign only sign

Extract 8.32 Teacher 1, Citizenship, first unit, class 1

1 TCH: Is that respect?
2 STU: No
3 TCH: It's not respecting. So we have to continue working very well to get a Sufi. Because all the time is a bien. Some people in the last class said 'I have a Bien' and I said you weren't respecting all during class…ahhh. So we have to be honest with respect.

In the case of Arts classes, *evaluation* is also the most frequent type of feedback (33.5%), followed by *meta-feedback* (26.3%), *expansion* (18.8%), and *re-route* (17.3%). When compared to Science lessons, a lot of significant differences appear (see **Table 8.14** below). *Evaluation* and *expansion* are more frequently deployed in Science, but *meta-feedback* is much more used in Arts. In Science, there is more *positive evaluation*, especially *with no recasts*, more *negative evaluation with explicit correction focusing on language* and more *didactic recasts*; in Arts, there is more *positive evaluation with recasts*, more *negative evaluation without explicit correction*, more *explicit correction focusing on content* and significantly more *prompts*.

	Science		Arts			
	%	N	%	N	*T*	χ^2
Evaluation	41.4%	1254	33.5%	182	3.49***	12.15***
Expansion	27.0%	817	18.8%	102	4.05***	16.35***
Revision	5.5%	166	4.2%	23	1.20	1.45
Re-route	20.2%	613	17.3%	94	1.60	2.55
Meta-feedback	5.9%	178	26.3%	143	15.85***	234.83***
Total		3028		544		
Evaluation						
Positive	88.7%	1109	80.1%	145	3.30***	10.81***
Negative	11.3%	141	19.9%	36	3.30***	10.81***
Total		1250		181		
Positive evaluation						
With recast	23.7%	261	32.9%	46	2.36***	5.56***
With no recast	76.3%	839	67.1%	94	2.36***	5.56***
Total		1100		140		
Negative evaluation						
With explicit correction	61.2%	85	38.9%	14	2.43***	5.77***
With no explicit correction	38.8%	54	61.1%	22	2.43***	5.77***
Total		139		36		
With explicit correction						
Content	45.9%	39	78.6%	11	2.30**	5.14**
Language	44.7%	38	7.1%	1	2.74***	7.10***
Content + Language	9.4%	8	14.3%	2	0.56	0.31
Total		85		14		
Re-route						
Prompt	79.0%	484	87.2%	82	1.87*	3.50*
Recast	21.0%	129	12.8%	12	1.87*	3.50*
Total		613		94		

Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$

Table 8.14 Comparison feedback types Science vs. Arts.

Extract 8.33 exemplifies the most frequent type of feedback in Science, *evaluation*. More specifically, this evaluation is a *negative evaluation with explicit correction focusing on content* ("Celts built round houses and not rectangular ones"). **Extract 8.34**, in turn, illustrates a very frequent type of feedback found in Arts lesson, *meta-feedback*. In it, the teacher is explaining the student how she can do better and she has to put more effort into the drawing, implying that she will not get a good mark.

Extract 8.33 Teacher 3, Science, second unit, class 1

1	TCH:	Cantabrian coast, north of the Peninsula. Instead of living in r- in rectangular houses, the Celts preferred to build
2	STU:	Square
3	TCH:	Round houses, ok? They were not rectangular, they were round houses, ok?

Extract 8.34 Teacher 3, Arts, first unit, classes 1 and 2

1	TCH:	<L1 Mira, Vanesa L1>, what do you get..? This is what you get.. ((Vanesa tries to reply)) No, no, no, listen to me. This is what you get when you talk a lot, when you don't pay attention, when you don't follow the rules, you don't behave, this is what you get, that some people have finished already and you are doing only the drawing. Vanesa, I give you a grade for effort and for the outcome, and this… you could have done it this ten times better, ok?

The most frequent type of feedback in Drama, as in all the other subjects, is *evaluation* (44.4%). *Expansion* (21.4%), *re-route* (18.3%), and *meta-feedback* (14.3%) are also present. In **Table 8.15**, differences between Science and Drama are illustrated. There is a weak significant difference regarding *revision*, more used in Science, and a stronger one for *meta-feedback*, which is more present in Drama classes. Weak significant differences are also found for *positive evaluation with or without recast*: *with recast* is more common in Drama, *without recast* in Science. All *explicit corrections* in Drama focus on *content*, in Science they are more balanced between *content and language*. In Science there are more *prompts* than in Drama, whereas in Drama there are more *recasts* than in Science. Finally, *recasts* in Science tend to focus on *language*, whereas in Drama there are more recasts focusing on *both content and language*.

	Science		Drama			
	%	N	%	N	*T*	χ^2
Evaluation	41.4%	1254	44.4%	56	0.68	0.46
Expansion	27.0%	817	21.4%	27	1.38	1.90
Revision	5.5%	166	1.6%	2	1.91*	3.64*
Re-route	20.2%	613	18.3%	23	0.55	0.30
Meta-feedback	5.9%	178	14.3%	18	3.84***	14.67***
Total		3028		126		
Positive evaluation						
With recast	23.7%	261	34.7%	17	1.75*	3.08*
With no recast	76.3%	839	65.3%	32	1.75*	3.08*
Total		1100		49		
With explicit correction						
Content	45.9%	39	100.0%	3	1.86*	3.40*
Language	44.7%	38	—	—	0.00	2.36
Content + Language	9.4%	8	—	—	0.00	0.31
Total		85		3		
Re-route						
Prompt	79.0%	484	60.9%	14	2.07**	4.27**
Recast	21.0%	129	39.1%	9	2.07**	4.27**
Total		613		23		
Recast						
Content	1.6%	2	—	—	0.00	0.14
Language	96.9%	125	77.8%	7	2.78***	7.40***
Content + Language	1.6%	2	22.2%	2	3.72***	12.77***
Total		129		9		
Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$						

Table 8.15 Comparison feedback types Science vs. Drama.

As *meta-feedback* is present in Drama (14.3%), **Extract 8.35** illustrates this kind of feedback in a Drama lesson. It cannot be discerned whether it is the genre of the subject or the teacher that influences the presence of *meta-feedback* in Drama, since the corpus includes Drama lessons only from one teacher (AfL). It occurred at the end of the lesson, in which the teacher made students aware of the good job they had done during the lesson.

Extract 8.35 Teacher 2 (AfL), Drama, second unit, class 1

1 TCH: I think that you've done very well this lesson, so well done. Ok, well done

Extract 8.36 shows examples of *re-route* in Science, which amount to 20.2%. Specifically, they are *prompts* (turns 5 and 7), as the teacher, through the means of repeating the same question several times, is trying to get students to arrive at the correct answer.

Extract 8.36 Teacher 3 (Non-AfL), Science, first unit, class 3

1	STU1:	amphibians stay in or near water to keep their skin dry. True
2	TCH:	Do you agree?
3	STU:	((Some)) no
4	STU:	Yes
5	TCH:	Is it true?
6	STU:	No!
7	TCH:	To keep their skin dry. Is it true or false, Gloria?
8	STU1	false

Citizenship and Arts are compared in **Table 8.16**. Many significant differences are found in the main categories: *evaluation*, *expansion*, and *re-route* happen to be more frequent in Citizenship, but *revision* and *meta-feedback* are more present in Arts. *Positive evaluation*, *positive evaluation with recasts focusing on language*, *negative evaluation with explicit correction* (especially *focusing on language*), *prompts which focus on content and language*, and *didactic recasts* are all significantly more frequent in Citizenship. In turn, *negative evaluation* in general, *negative evaluation without explicit correction*, *negative evaluation with explicit correction focusing on content*, and *prompts* are significantly more numerous in Arts lessons.

	Citizenship		Arts			
	%	N	%	N	*T*	χ^2
Evaluation	44.4%	1049	33.5%	182	4.68***	21.74***
Expansion	24.2%	571	18.8%	102	2.71***	7.31***
Revision	0.8%	20	4.2%	23	5.92***	34.68***
Re-route	20.2%	477	17.3%	94	1.54	2.38
Meta-feedback	10.4%	245	26.3%	143	10.00***	96.80***
Total		2362		544		
Evaluation						
Positive	89.5%	938	80.1%	145	3.62***	13.01***
Negative	10.5%	110	19.9%	36	3.62***	13.01***
Total		1048		181		
With recast						
Content	2.4%	6	6.5%	3	1.49	2.22
Language	94.0%	234	87.0%	40	1.70*	2.89*
Content + Language	3.6%	9	6.5%	3	0.92	0.84
Total		249		46		
Negative evaluation						
With explicit correction	71.6%	78	38.9%	14	3.67***	12.46***
With no explicit correction	28.4%	31	61.1%	22	3.67***	12.46***
Total		109		36		
With explicit correction						
Content	33.8%	26	78.6%	11	3.29***	9.86***
Language	48.1%	37	7.1%	1	2.96***	8.15***
Content + Language	18.2%	14	14.3%	2	0.35	0.12
Total		77		14		
Re-route						
Prompt	73.4%	350	87.2%	82	2.88***	8.19***
Recast	26.6%	127	12.8%	12	2.88***	8.19***
Total		477		94		
Prompt						
Content	47.4%	166	58.5%	48	1.81*	3.28*
Language	21.1%	74	23.2%	19	0.40	0.16
Content + Language	31.4%	110	18.3%	15	2.37***	5.57***
Total		350		82		

Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$

Table 8.16 Comparison feedback types Citizenship vs. Arts.

Expansion is a frequent feedback type in Citizenship, and so it is illustrated in **Extract 8.37**: in turns 2 and 4, the teacher tries to expand on students' thinking through the use of meta-cognitive questions. **Extract 8.38**, in turn, is showing an example of *revision* in an Arts lesson, since Arts and Science are the two subjects in which *revision* is more present.

Extract 8.37 Teacher 2 (AfL), Citizenship, second unit, classes 1 and 2

1 STU1: About the health
2 TCH: About being healthy, good. Um, why are we looking at being healthy? Mateo?
3 STU2: To have a good diet and…to be good.
4 TCH: ((correction)) To be good, to feel good. Good, why else do you think, Bárbara?
5 STU3: to know how eh, we are eh, how is the level of our health.

Extract 8.38 Teacher 3 (Non-AfL), Arts, second unit, class 1

1 TCH: So.. they focused on elements that belonged to mass culture, to the culture of people in general, ok? Things that were very popular, popular at that time, including personalities, ok?

In **Table 8.17**, feedback types in Citizenship and Drama are compared. The only significant differences between the two are that *positive evaluation with recast* tends to focus more on *content* in Drama classes and also that all *explicit corrections* focus on *content* in Drama, while in Citizenship the focus is more balanced.

	Citizenship		Drama			
	%	N	%	N	*T*	χ^2
Evaluation	44.4%	1049	44.4%	56	0.01	0.00
Expansion	24.2%	571	21.4%	27	0.70	0.49
Revision	0.8%	20	1.6%	2	0.86	0.75
Re-route	20.2%	477	18.3%	23	0.53	0.28
Meta-feedback	10.4%	245	14.3%	18	1.39	1.94
Total		2362		126		
With recast						
Content	2.4%	6	11.8%	2	2.20**	4.77**
Language	94.0%	234	88.2%	15	0.93	0.88
Content + Language	3.6%	9	—	—	0.00	0.64
Total		249		17		
With explicit correction						
Content	33.8%	26	100.0%	3	2.40***	5.48***
Language	48.1%	37	—	—	0.00	2.68
Content + Language	18.2%	14	—	—	0.00	0.66
Total		77		3		
Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$						

Table 8.17 Comparison feedback types Citizenship vs. Drama.

As an illustration of *negative evaluation with explicit correction on content* in a Drama lesson, see **Extract 8.39** below (turns 5 and 7). The teacher is correcting students, since she is not asking for the acting out of a particular person, but rather the acting out of a character type.

Extract 8.39 Teacher 2 (AfL), Drama, second unit, class 1

1 TCH: So doing all the different things? Ok, we'll do one more….ok, who would like to choose a character? Who would like to choose a character? ((students raise hands)) ok, everybody listen to Laura.
2 STU1: Eh, ok…..I have to….ahh! Ok, the %L1….biblotecaria….L1%
3 TCH: Librarian
4 STU: Yeahhhh!!
5 TCH: No, no, no, no, no….stop, stop, stop….no, not acting out a person….a character type.
6 STU: Eh, but the person that….%x….x%
7 TCH: In general….we're not talking about a person that you know. A character type.

Types of feedback in Arts and Drama lessons are compared in **Table 8.18**. The few significant differences that were found are the following: *evaluation* and *didactic recasts* are more recurrent in Drama, while *meta-feedback* and *prompts* are more common in Arts.

	Arts		Drama			
	%	N	%	N	*T*	χ^2
Evaluation						
Expansion	33.5%	182	44.4%	56	2.33**	5.39**
Revision	18.8%	102	21.4%	27	0.69	0.47
Re-route	4.2%	23	1.6%	2	1.41	1.99
Meta-feedback	17.3%	94	18.3%	23	0.26***	0.07***
Total	26.3%	143	14.3%	18	2.85	8.07
Re-route						
Prompt	87.2%	82	60.9%	14	3.04***	8.72***
Recast	12.8%	12	39.1%	9	3.04***	8.72***
Total		94		23		
Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$						

Table 8.18 Comparison feedback types Arts vs. Drama.

In **Extract 8.40**, Teacher 3 is giving students *prompts* to help them arrive at the right response (turns 4, 6 and 8). This leads to one student's successful answer in the final turn of the extract.

Extract 8.40 Teacher 3 (Non-AfL), Arts, second unit, class 1

1	TCH:	where do you think the word pop comes from?
2	STU:	From music
3	STU:	From the music
4	TCH:	Yeah, but pop music? Pop. Do you-
5	STU:	Modern
6	TCH:	Modern?
7	STU:	Modern pop
8	TCH:	And pop comes from?
9	STU:	((Some)) [popular]
10	TCH:	[The short], the short form of?
11	STU:	((Some)) popular

Extract 8.41 shows an example of *evaluative* feedback in a Drama lesson, more specifically *positive evaluation with recast* in turn 3.

Extract 8.41 Teacher 2 (AfL), Drama, second unit, class 1

1	TCH:	Once you have the events, what did you have to do with those events? Ana?
2	STU:	The scene
3	TCH:	Good, put those events into different scenes. Next step, what did you do next? Miguel?

Results from this section have shown that, in spite of some common findings across subjects, such as the fact that *evaluation* and *expansion* are the most frequent type of feedback while *revision* is very scarce, significant differences regarding the use of feedback types appear across subjects. The only exception is the comparison between Citizenship and Drama, in which significant differences are very few. In the other comparisons, significant differences across subjects appear in both the main feedback types and their subtypes. The presence of *meta-feedback* is especially frequent in Arts and Drama lessons, which may indicate that there is more room for reflection about learning and assessment in these two subjects. It is important to take into account that, with the exception of Drama, the other subjects were taught by two (Science and Arts) or three (Citizenship) different teachers, and thus, the results might be related to teachers' styles rather than to type of subject. It is, therefore, necessary to examine the types of feedback used by each teacher in different subjects.

8.5.2 Feedback type depending on teacher and subject

This subsection will analyse differences in feedback types used by teachers in the two different subjects they teach. Teacher 1, who is an AfL teacher, teaches Science and Citizenship; Teacher 2, also AfL, teaches Citizenship and Drama; Teacher 3, a Non-AfL teacher, teaches Science and Arts; and, finally, Teacher 4, also Non-AfL, teaches Citizenship and Arts. Results will present feedback types used by each of the teachers in their two subjects. Thus, any significant differences that may appear will be due to the subject and not to the teacher. In this way, we address the necessity of checking what variable is more influential as to the type of questions asked, subject or teacher.

In **Table 8.19**, we can see that Teacher 1 uses significantly more *expansion*, *revision*, *prompts*, and *prompts focusing on language* in Science. In Citizenship, however, Teacher 1 uses more *meta-feedback*; more *positive evaluation with recasts*, and these focusing on *language*; *negative evaluation with explicit correction focusing on content* more frequently than in Science; and more *didactic recasts* in all their subtypes. To see illustrations of Teacher 1 using *expansion* in Science and *meta-feedback* in Citizenship, see **Extracts 8.31 and 8.32** above, respectively.

	Science		Citizenship			
	%	N	%	N	*T*	χ^2
Evaluation	40.6%	687	43.6%	488	1.58	2.50
Expansion	29.0%	490	22.0%	246	4.13***	16.96***
Revision	2.2%	38	—	—	0.00	25.48***
Re-route	18.1%	306	17.6%	197	0.32	0.11
Meta-feedback	10.0%	170	16.8%	188	5.28***	27.64***
Total		1692		1119		
Positive evaluation						
With recast	7.4%	125	10.4%	116	2.76***	7.62***
With no recast	27.8%	470	26.8%	300	0.56	0.32
Total		1692		1119		
With recast						
Content	0.3%	5	0.2%	2	0.61	0.37
Language	6.7%	114	9.8%	110	2.97***	8.78***
Content + Language	0.4%	6	0.4%	4	0.01	0.00
Total		1692		1119		
Negative evaluation						
With explicit correction	2.9%	49	4.5%	50	2.21**	4.90**
With no explicit correction	2.0%	33	1.5%	17	0.85	0.72
Total		1692		1119		
With explicit correction						
Content	0.6%	10	1.7%	19	2.85***	8.08***
Language	1.9%	32	2.1%	23	0.31	0.09
Content + Language	0.4%	7	0.6%	7	0.78	0.61
Total		1692		1119		
Re-route						
Prompt	15.1%	256	12.6%	141	1.89*	3.55*
Recast	3.0%	50	5.0%	56	2.79***	7.80***
Total		1692		1119		
Prompt						
Content	6.0%	101	6.3%	71	0.41	0.17
Language	5.8%	98	2.7%	30	3.88***	15.00***
Content + Language	3.4%	57	3.6%	40	0.29	0.09
Total		1692		1119		
Recast						
Content	0.1%	1	—	—	0.00	0.66
Language	2.8%	48	4.6%	52	2.54***	6.43***
Content + Language	0.1%	1	0.4%	4	1.84*	3.38*
Total		1692		1119		

Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$

Table 8.19 Comparison feedback types for Teacher 1.

Teacher 2's types of feedback in Citizenship and Drama are illustrated in **Table 8.20**. Few significant differences are found across the two subjects. In Citizenship, Teacher 2 uses *expansion* more frequently. In Drama, she uses significantly more *meta-feedback* and *negative evaluation*, and specifically *negative evaluation without explicit correction*; she also uses more *explicit correction focusing on content* in Drama than in Citizenship lessons. Again, Teacher 2's different uses of feedback are illustrated above: **Extract 8.37** illustrates Teacher 2 using *expansion* in a Citizenship lesson while **Extract 8.35** shows how Teacher 2 used *meta-feedback* in a Drama lesson.

	Citizenship		Drama			
	%	N	%	N	*T*	χ^2
Evaluation	44.8%	162	44.4%	56	0.06	0.00
Expansion	31.2%	113	21.4%	27	2.10**	4.38**
Revision	0.6%	2	1.6%	2	1.11	1.23
Re-route	16.6%	60	18.3%	23	0.43	0.19
Meta-feedback	6.9%	25	14.3%	18	2.53***	6.33***
Total		362		126		
Evaluation						
Positive	43.6%	158	38.9%	49	0.93	0.87
Negative	1.1%	4	5.6%	7	2.92***	8.40***
Total		362		126		
Negative evaluation						
With explicit correction	1.1%	4	2.4%	3	1.04	1.08
With no explicit correction	—	—	3.2%	4	0.00	11.59***
Total		362		126		
With explicit correction						
Content	—	—	2.4%	3	0.00	8.67***
Language	0.3%	1	—	—	0.00	0.35
Content + Language	0.8%	3	—	—	0.00	1.05
Total		362		126		

Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$

Table 8.20 Comparison feedback types for Teacher 2.

Similarly to Teacher 1, Teacher 3 also uses feedback types differently in her two subjects, Science and Arts (see **Table 8.21**). In Science, *evaluation*, *expansion*, *revision*, and *re-route* are more frequent; in Arts, *meta-feedback* is. In Science, Teacher 3 also uses more *positive evaluation*, and specifically *positive evaluation with no recast*; *negative evaluation with explicit correction* is more used in Science too, as well as *explicit corrections focusing on content*; *didactic recasts* in general and *didactic recasts focusing on language* are more frequent in Science; and, finally, *prompts* which combine focus on *content and language* are, again, more numerous in Science. In Arts, apart from *meta-feedback*, Teacher 3 uses more *explicit correction focusing on both content and language*. As an illustration of Teacher 3's types of feedback in Science and Arts, see **Extracts 8.33 and 8.34** above: **Extract 8.33** illustrates the use of *negative evaluation with explicit correction focusing on content* in Science, whereas **Extract 8.34** exemplifies *meta-feedback* in an Arts lesson.

	Science		Arts			
	%	N	%	N	*T*	χ^2
Evaluation	42.4%	567	30.1%	103	4.16***	17.16***
Expansion	24.5%	327	17.8%	61	2.60***	6.72***
Revision	9.6%	128	6.7%	23	1.64	2.70
Re-route	23.0%	307	16.7%	57	2.52***	6.36***
Meta-feedback	0.6%	8	28.7%	98	21.49***	362.45***
Total		1337		342		
Evaluation						
Positive	37.9%	507	27.2%	93	3.71***	13.65***
Negative	4.4%	59	2.6%	9	1.49	2.22
Total		1337		342		
Positive evaluation						
With recast	10.2%	136	10.5%	36	0.19	0.04
With no recast	27.6%	369	15.5%	53	4.63***	21.20***
Total		1337		342		
Negative evaluation						
With explicit correction	2.7%	36	0.9%	3	1.99**	3.96**
With no explicit correction	1.6%	21	1.8%	6	0.24	0.06
Total		1337		342		
With explicit correction						
Content	2.2%	29	0.3%	1	2.34**	5.47***
Language	0.4%	6	—	—	0.00	1.54
Content + Language	0.1%	1	0.6%	2	1.99**	3.97**
Total		1337		342		
Re-route						
Prompt	17.1%	228	14.6%	50	1.08	1.17
Recast	5.9%	79	2.0%	7	2.90***	8.36***
Total		1337		342		
Prompt						
Content	10.5%	141	9.4%	32	0.65	0.42
Language	2.6%	35	3.2%	11	0.60	0.37
Content + Language	3.9%	52	2.0%	7	1.65	2.73*
Total		1337		342		
Recast						
Content	0.1%	1	—	—	0.00	0.26
Language	5.8%	77	1.8%	6	3.06***	9.30***
Content + Language	0.1%	1	0.3%	1	1.04	1.08
Total		1337		342		

Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$

Table 8.21 Comparison feedback type for Teacher 3.

As it occurred with the other teachers, Teacher 4 used feedback types differently in his two subjects (see **Table 8.22**). *Revision* and *re-route* were more used in Citizenship, while *meta-feedback* is much more frequent in her Arts lessons. In Citizenship, this teacher uses more *didactic recasts*, more *positive evaluation* in general, as well as *with and without recast*, and *positive evaluation with recast focusing on language*. In turn, *negative evaluation* in general, and particularly, *with and without explicit correction* are more used by Teacher 4 in Arts classes, and *explicit corrections that focus on content* are also more significant in this subject.

	Citizenship		Arts			
	%	N	%	N	*T*	χ^2
Evaluation	45.3%	399	39.1%	79	1.60	2.55
Expansion	24.1%	212	20.3%	41	1.14	1.30
Revision	2.0%	18	—	—	0.00	4.20**
Re-route	25.0%	220	18.3%	37	2.01**	4.02**
Meta-feedback	3.6%	32	22.3%	45	9.69***	86.50***
Total		881		202		
Evaluation						
Positive	40.9%	360	25.7%	52	4.02***	15.94***
Negative	4.4%	39	13.4%	27	4.84***	22.95***
Total		881		202		
Positive evaluation						
With recast	8.7%	77	5.0%	10	1.79*	3.19*
With no recast	31.6%	278	20.3%	41	3.18***	10.02***
Total		881		202		
With recast						
Content	0.3%	3	1.0%	2	1.23	1.51
Language	8.1%	71	3.5%	7	2.28**	5.19**
Content + Language	0.3%	3	0.5%	1	0.33	0.11
Total		881		202		
Negative evaluation						
With explicit correction	2.7%	24	5.4%	11	1.97**	3.89**
With no explicit correction	1.6%	14	7.9%	16	5.00***	24.46***
Total		881		202		
With explicit correction						
Content	0.8%	7	5.0%	10	4.32***	18.37***
Language	1.5%	13	0.5%	1	1.11	1.24
Content + Language	0.5%	4	—	—	0.00	0.92
Total		881		202		
Re-route						
Prompt	19.8%	174	15.8%	32	1.28	1.63
Recast	5.2%	46	2.5%	5	1.66*	2.76*
Total		881		202		

Note: *$p < 0.01$; **$p < 0.05$; ***$p < 0.001$

Table 8.22 Comparison feedback types for Teacher 4.

Extract 8.42 shows an example of *re-route* in one of Teacher 4's Citizenship lessons. In turn 3, the teacher tries to make students explain what rights are without translating the word into Spanish.

Extract 8.42 Teacher 4, Citizenship, second unit, class 1

1 TCH: Ok, and the rights, tell me the rights?
2 STU: %L1…Derechos…L1%
3 TCH: %x…x% but what are rights? Explain to me because I don't understand.
4 STU: That you have to…..put..

Extract 8.43 exemplifies a *negative evaluation with explicit correction on language*, used by Teacher 4 in one of his Arts lessons.

Extract 8.43 Teacher 4, Arts, first unit, class 1

1 TCH: Orange, ok, red..%x…x% an building
2 STU: What?
3 TCH: You say pardon, ok, "pardon?", when you don't understand something, ok we don't say "what" anymore, we say pardon

In general, then, results in this subsection have shown how the variable subject is also important when it comes to types of feedback used by teacher, as the same teacher in different subjects uses some of the feedback types significantly different. The only case in which there were not many differences was in the case of Teacher 2. This fact may mean that Teacher 2's AfL background plays a bigger role than the specific pedagogy of the subjects she teaches.

8.6 FEEDBACK TYPES AND STUDENT UPTAKE

The present section will present a qualitative account of students' uptake after feedback. Uptake after third moves can play a crucial role for learning (Barnes 1975). Research has shown that for uptake to occur and for co-construction of meaning and opportunities for new learning cycles to appear, third moves need to be follow-up/feedback moves rather than evaluating ones (Wells 1993; Nassaji & Wells 2000; Hall 1998; Nystrand 1997; Rex & McEachen 1999; Cortés-Conde 2000; Boyd & Maloof 2000; Consolo 2000; Duff 2000; Sullivan 2000).

As we have already seen in some previous extracts and, as **Extract 8.44** shows, when there was *evaluation* in the teacher's feedback turn, uptake was not likely to occur.

Extract 8.44 Teacher 1, Science, first unit, class 2

1 TCH: What is there?
2 STU1: the table.. of contents
3 TCH: The table of contents. What's in the table of contents ((al students start answering at the same time)). What is it, Bea?
4 STU2: in the table of contents there is what we <x…x> in the book and the page that

As shown in the extract, after the teacher *positive evaluation* in turn 3, there is no uptake on the students' part, but rather there is topic continuation, the teacher still has the floor and goes on making questions.

As illustrated in **Extract 8.45**, when the *evaluation* was *negative with explicit correction*, there was, again, topic continuation on the teacher's part (in this case, with expansion) with no further student uptake.

Extract 8.45 Teacher 3, Science, second unit, class 1

1	TCH:	Cantabrian coast, north of the Peninsula. Instead of living in r- in rectangular houses, the Celts preferred to build
2	STU:	Square
3	TCH:	Round houses, ok? They were not rectangular, they were round houses, ok? And a difference between the Iberian and the Celts is that these people knew how to use metals. So they eh.. built or they created a lot of tools and weapons using metal, ok?

When *negative evaluation* did *not include explicit correction*, the teacher usually continued with some kind of *re-route*, and then, student uptake took place. **Extract 8.46** perfectly illustrates this case: in turn 3, the teacher negatively evaluates the previous student response ("No, it isn't right"). Then, she makes use of *re-route*, giving the student clues as to why the previous response was incorrect ("so look at my first word and look at your first word"). This leads to successful student uptake.

Extract 8.46 Teacher 1, Citizenship, first unit, class 3

1	TCH:	Is this right?
2	STU:	Yes
3	TCH:	No, it isn't right. So look at my first word and look at your first word. Ok
4	STU:	You have to put the comma
5	TCH:	Thank you. Álvaro is paying attention. I said the first word

In many occasions, when there was *expansion* in the teacher's feedback move, there was student uptake. In most of these cases, this occurred when the *expansion* was done through a question, as illustrated in **Extract 8.47**. In turn 5, there is *expansion* on the previous question and answer, through a why-question, which is taken up by the student in turn 6 with a quite long intervention.

Extract 8.47 Teacher 2, Citizenship, first unit, classes 1 and 2

1	TCH:	You know, when I first saw it, I thought he was eating ice cream. Say if it was rice, would that be healthy or not?
2	STU:	((All)) Yes
3	TCH:	And if it was ice cream?
4	STU:	((All)) No
5	TCH:	Why is it not so healthy to eat ice cream? Bárbara?
6	STU:	Because it's like a bomb to your stomach. Because it's so, em, cold

However, not all *expansions* are questions, and these may not be taken up by students. For instance, in **Extract 8.48**, students are giving examples of rights. In turn 3, there is *expansion*, as the teacher offers more information related to the exchange, trying to explain the student why his example is not valid. However, after this *expansion*, the next IRF starts, as he asks the next student to give another example.

Extract 8.48 Teacher 4, Citizenship, second unit, class 5

1	TCH:	But do you get money for your work?
2	STU:	No
3	TCH:	So, the only thing you cannot pay for your... for your meals, yes or no? For the food, OK? So you have the right for someone else to pay for your lunch, yes or no? OK? Good. Cristian.
4	STU1:	the babies don't have the right to walk.

Generally, when there is *revision* in the feedback move, no student uptake follows, as the teacher then asks a question that helps students continue with the review (as in **Extract 8.49**), or a question that starts a new episode in the lesson.

Extract 8.49 Teacher 3, Arts, second unit, class 1

1	TCH:	So.. they focused on elements that belonged to mass culture, to the culture of people in general, ok? Things that were very popular, popular at that time, including personalities, ok? What type of colours did they use? Diane ((she doesn't answer)) What type of colours did they use..?
2	STU:	bright colours

Whenever there is *re-route*, there is learner uptake in most of the cases. This is to be expected, as *re-route* aims at offering learners scaffolding and assistance to arrive to the right answer. In **Extract 8.50**, we find quite a long extract in which the teacher has to make use of *re-route* several times (turns 3, 7, 11). The students do not seem to understand that the teacher is asking about the main idea of the first paragraph, and not about the main idea of the whole text. To make students understand what he means and what he wants from them, he has to give them clues and offer them help. After every teacher's *re-route* move, student uptake occurs, since, in addition, *re-route* feedback moves are, at the same time, questions: in turn 4 the student just manages to provide a trun*cated response*, in turn 8 another student offers *one-phrase response*, and finally, the answer given by the student in turn 12 (who is the same student as in turn 4) is the correct one and is made up of *one clause*.

Extract 8.50 Teacher 4, Citizenship, second unit, class 4

1	TCH:	For example, in this paragraph number one, who told me the main idea? Aitor, what do you think?
2	STU1:	Equality between the woman and….
3	TCH:	Yeah, this is for the whole text, but what is the idea of the first paragraph? That you read? Carlos?
4	STU2:	That eh, mmm, ((gestures that he doesn't know))
5	TCH:	Cristian?
6	STU3:	Eh, that em, educational levels of the man and the woman have to be the same.
7	TCH:	Ok, but in this paragraph? You get this? No, that wasn't in that paragraph. Which paragraph did Aitor read?
8	STU:	The first
9	TCH:	The first, ok? Lucia?
10	STU4:	the man and the woman have to be equal, eh, equality
11	TCH:	Equal ((correction)) Mmm, ok, but this is the general idea, but what is the idea in the first paragraph? The thing that right now Aitor read, ok? What is the first line? Carlos, please.
12	STU2:	((reading)) that equality between men and woman is fundamental.

Meta-feedback was not normally followed by student uptake. In general, *meta-feedback* was meant to make students reflect on learning or assessment at a personal level, as shown in **Extract 8.51**, in which students have to self-assess themselves.

Extract 8.51 Teacher 1, Citizenship, first unit, class 3

1	TCH:	No, it's very easy to say no, no, "notable, no or sobresaliente", no you have to ((points to poster)) look at the paper. So the first thing is respect, and respect to get a bien is to "work respectfully in class all of the time" and some people can't say that. The maximum for some people here is a sufi.
2		((class silent))
3	TCH:	Right. Right or wrong?
4	STU:	Right
5	TCH:	So let's try to get a sufi plus I'm going to try more in 30 minutes

When the reflection on learning or assessment was meant to be aloud, then teachers used *meta-questions* (see turn 1 in **Extract 8.52**). Also, these *meta-questions* were many times followed by *meta-feedback* (turns 3 and 5 from **Extract 8.52**). In **Extract 8.52**, the teacher is making the student reflect on her peer-assessment, making her see why it was not acceptable, and how to be more responsible when it comes to peer- or self-assessment. In this way, both her processes of learning and assessment will be positively affected in the future.

Extract 8.52 Teacher 1, Science, second unit, class 3

1	TCH:	What did you do well in your paper?
2	STU1:	I do very well this homework because I put four or five things but I put that they didn't have nothing from the house and then I put that almost have a <L1 notable L1> or a <L1 sobre L1>, and you put.. that it's impossible.
3	TCH:	Ok, so my problem with Mmm- with María is her idea. Eh.. Bea is a good friend of mine and she spoke clearly and loudly, so one <L1 sobre L1> for her
4	STU:	Ha!
5	TCH:	No, because they were certain things that you had to do for a <L1 sobre L1> or a <L1 notable L1> in the presentation. So you have to be more realistic. When I say this is how you get a <L1 sobre L1>, or a <L1 notable L1> or a <L bien L>, then you just can give eh.. a big present. Many people memorised the presentation.

This section has described, in a qualitative manner, when and why there is student uptake after certain types of feedback and not after other types. The feedback types that obtain higher student uptake are *expansion* and *re-route*. After *evaluation* and *revision*, it is very rare to find student uptake. After *meta-feedback*, more important than posterior student uptake is the inner reflection that the teacher encourages students to make regarding learning and assessment through the use of his/her *meta-questions* and through *meta-feedback*.

8.7 DISCUSSION

This chapter has reported on teachers' feedback. The third move of IRF sequence has been extensively researched, and its importance and complexity vastly highlighted (Nassaji & Wells 2000; Wells 1993; van Lier 2000; Mortimer & Scott 2003; Lyster 2007; Alexander 2004 are just a few examples). This chapter has presented the types of feedback used by the four teachers in our study, comparing and showing similarities and differences between the AfL and the Non-AfL ones. This chapter has also shown that certain types of feedback are more likely to follow certain types of questions and not others. In addition, types of feedback have been also studied taking into account the previous student response, the different subjects, the same teacher in two different subjects, and the student uptake (or lack of it) following feedback.

The quality of feedback is a key feature for AfL, since different studies have proven it enhances learning (Black & Wiliam 1998a: 36; Davison & Leung 2009; Harrison & Howard 2009). The analysis of the primary CLIL teachers' feedback in this study has shown that *evaluation* is the most frequent type, as shown by previous studies. As claimed by different researchers (Nassaji & Wells 2000; Wells 1993; Zhang Waring 2008; Llinares & Morton 2010), whenever there is *evaluation* in the third move, students' participation is limited, and when *evaluation* is avoided, interaction gets more equal. In general, *positive evaluation* has been found to be more frequent

than *negative*, often without including *recasts*, and when these are used, these *recasts* normally focus on *language*. *Negative evaluation* is normally accompanied by *explicit correction*, and what is focused on varies across teachers (content, language, or both).

Although *evaluation* can be formulaic for certain teachers (Mercer 2000; Alexander 2004), *evaluation with recast* may alter and enrich student's response, and place it in a more relevant context in terms of content, language, or both, thus being retroactively contextualizing (Lemke 1982). In spite of this, in the present study, *positive evaluations without recast* are much more common than positive evaluations *with recast*, regardless of the type of teacher.

Turning now to *re-route*, the study showed that, as *evaluation*, this type of feedback was also frequent in the CLIL primary data analysed. Let us remember that *re-route* is used when the student has provided an incorrect answer in the second turn, and so the teacher decides to help the student so that s/he can deal with the problem. This type of move is also known as scaffolding, a strategy which has been said to help integrating content and language (Lyster 2007), and has proved to have positive effects at the students' cognitive and meta-cognitive levels (van de Pol et al. 2010). This means that this type of feedback can help students to advance in their understanding of the content, and to improve the learning process. Regarding the subtypes of *re-route*, *prompts* were more frequent than *recasts* for all teachers, which is in contradiction with Llinares and Lyster (2014). This difference may be explained by the fact that, in the latter study, only *language recasts* were the focus, whereas in this study all *recasts* were taken into account (recasts focusing on language, on content, or on both aspects combined). Many researchers have shown a preference for *prompts* over *recasts* (Lyster & Ranta 1997; Ellis et al. 2001a, 2001b; Ellis & Sheen 2006): *recasts* lead to a repair consisting in repetition, which cannot be considered as evidence of learning, whereas with the use of *prompts*, there are more opportunities for learning.

The second centre of attention of the chapter was the comparison of the use of feedback types by AfL teachers and Non-AfL teachers. Similarly to results in previous chapters, significant differences also appear when feedback types are compared in these two types of schools: *expansion* and *meta-feedback* are the two types of feedback that AfL teachers use significantly more often, whereas Non-AfL teachers use *re-route* and *revision* more frequently. AfL teachers were expected to use more *expansion* and *meta-feedback* because both types of feedback fit very well with an AfL approach. As *re-route* moves are contingent and responsive, AfL teachers were expected to use them more frequently. However, Non-AfL teachers used this type of feedback significantly more often. This can be explained because, despite not being trained in AfL pedagogy, AfL is about quality teaching and learning, and so some AfL techniques can be

used by these teachers as well. As for *evaluation*, no significant differences appear, although it would have been expected that Non-AfL teachers would use it more frequently, since IRE would be more aligned to a traditional pedagogy and less aligned with an AfL pedagogy. Regarding *explicit correction*, it was expected to be infrequent in AfL classes, as it has been confirmed, even though it is even a bit more infrequent in Non-AfL classes. These results may be explained because all the classes are CLIL classes, which are essentially focused on students' understanding and participation.

In the light of these results, another of our initial hypothesis is verified: as it occurred with questions and responses when the two types of schools were compared, here we also find significant differences between the types of feedback deployed by teachers in AfL and Non-AfL schools. Consequently, one could say that, even though *evaluation* is still the most frequent feedback type, AfL teachers use third turns not only to evaluate but also to expand on students' understanding and their process of learning (Alexander 2004; Nystrand et al. 1997). Non-AfL teachers, with their use of *revision* and *re-route*, and not *evaluating* more frequently than AfL teachers, seem to be aligning with AfL pedagogy as well. As said elsewhere, CLIL methodology is about student participation and understanding. Whether the third turn *evaluates*, or is used to *extend* students' answers, to make connections, to *re-route*, to offer feedback on the learning process… has serious consequences for learning (Barnes 1975; Lyster 1998). When *evaluation* is avoided, and *expansion* or *re-route* are used instead, IRF sequences can be *extended*, the feedback move being followed by another student's response, which again received feedback and so on (Mortimer & Scott 2003; van Lier 1988). On the contrary, when too much *evaluation* is used exchanges are very short (and so are student responses), there is no co-construction of meaning, and expanding students' reasoning, thinking, and learning is not the main goal. Although learning has not been measured, this study may lead us to think that learning processes are improved when long and engaging IRF sequences (which are more common in AfL classrooms) are created through the use of different types of feedback which are not mere *evaluation*.

To finish with the comparison of AfL and Non-AfL teachers, let us consider the use of *meta-feedback*. AfL teachers, as we have seen, use *meta-feedback* significantly more often than Non-AfL teachers (12.2% as opposed to 6.6%). *Meta-feedback* has been proved to be a distinctive feature of AfL discourse. Many authors have stressed the importance of focusing on the learning process through the identification of students' strengths and weaknesses (Black & Wiliam 1998b; Leung 2007; Harrison & Howard 2009; Black et al. 2003; Black et al. 2004; Harlen & Winter 2004). The present study has demonstrated how this is achieved through *meta-questions* and *meta-feedback*.

A third focus of this chapter has been the relationship between question type and feedback type. The results have shown that *questions for facts* and *language questions* receive significantly more *evaluations* (see also Nassaji & Wells 2000); *meta-questions* are often followed by *meta-feedback*; and *expansion* is mainly found after *questions for reasons* (see also Nassaji & Wells 2000) and *meta-cognitive* questions. *Questions for opinions* are also followed by *evaluation*, contrary to what may be expected in an environment of inquiry. These results echo Nassaji and Wells' (2000), in which negotiatory questions gave way to evaluating feedback moves. As a way of explaining why some negotiatory questions still receive a lot of *evaluations* (as it is the case in the present study, too), they argue that the teacher may repeat the response for all students to pay attention to it and as a starting point for further elaboration (Nassaji & Wells 2000: 397). Nonetheless, in this study, *questions for opinions* were not always followed by *evaluation* but also by *expansion*.

When response and feedback were analysed in combination, it was very interesting to discover that most of *minimal responses* were followed by *expansion*, where the teacher tried to get the students to expand on their thinking. This *expansion* (used very frequently by AfL teachers) was frequently another initiation at the same time: a "why" or "how" question. *Re-route*, present in the discourse of all teachers, was also found to be an initiation at the same time that a feedback move. According to Wells (1993), third turns which are, at the same time, feedback and the initiating move of the next exchange have positive effects, as they force students to take responsibility for what they have previously said, and if they are "why" or "how questions, they help redirect and challenge students' thinking, providing more information about their understandings (Ruiz-Primo & Furtak 2007). *Truncated responses*, on their part, were very frequently followed by *re-route*, as teachers tried to guide students through *prompts* or *recasts* to the correct response they were struggling to produce. This is in line with studies which claim that incorrect responses can lead to opportunities for learning if the feedback provided helps the student find the source of their misunderstanding (Hattie & Timperley 2007; Zhang Waring 2008; Harrison & Howard 2009). This investigation has also shown that *evaluation* follows short and simple responses as well as long and more complex ones, thus contradicting Wells (1993) and Nassaji and Wells (2000), who claim that when there is no *evaluation*, students' contributions are longer and more complex.

The next focus of the chapter was the analysis of feedback in the different subjects, which has also rendered significant differences. It is important to remember that the two AfL teachers teach Science, Citizenship and Drama, and the two Non-AfL teachers teach Science, Citizenship and Arts. Except for Drama, there are data of different teachers teaching the same subject. *Evaluation* is more frequently found in Citizenship and Drama; *expansion* is more present in

Science (followed by Arts); *revision* in Science and Arts; *re-route* presents no significant differences between subjects, and finally, *meta-feedback* is more frequently used in Arts (followed by Drama). It seems that *expansion* and its frequent use in Science is not due to the teacher but to the genre, as we have one teacher of each type in this subject. It may be explained by the fact that students are faced with complex scientific phenomena, and so the teacher needs to use more *expansion* to ensure understanding. The same can be said about *meta-feedback*, which is significantly more used by AfL teachers but more frequent in Arts (with no AfL teachers teaching it). An explanation to this finding is that in Arts teachers ask to evaluate one's and each other's work (Dale & Tanner 2012), with the teacher providing arguments about what has been done well and what needs further improvement. An alternative explanation is that the activities done in the Arts lessons were hands-on, and students asked the teacher whether their work was going well or something needed to be changed. *Evaluation* is very frequent in general for all teachers and subjects. *Revision* being more frequent in Science may not be affected by the teacher, since there are both an AfL and a Non-AfL teacher teaching it, but Arts may be, as the two teachers teaching Arts are the two teachers using *revision* more frequently.

When each teacher is compared teaching their two subjects, we find that, in general, to a lesser or greater extent, teachers change the way in which they use feedback (regardless of whether they are AfL or Non-AfL), which reinforces the finding of different subjects playing their role. Teacher 1 uses more *expansion* and *revision* in Science, and more *meta-feedback* in Citizenship. Teacher 2 deploys more *expansion* in Citizenship and more *meta-feedback* in Drama. In Science, Teacher 3 uses more *evaluation*, *expansion* and *re-route*, and more *meta-feedback* in Arts. Teacher 4 uses differently *revision*, *re-route* (more in Citizenship), and *meta-feedback* (more in Drama). These results coincide with results of feedback types depending on the subject: *revision* is significantly more present in Science than in other subjects, and *meta-feedback* in Arts. *Meta-feedback* is the feedback type which changes for all teachers, *expansion* changes for both AfL teachers, and *re-route* changes for Non-AfL teachers. All this comes to confirm one of our initial hypotheses: different types of feedback are found when teachers teach different subjects (the same occurred with questions and responses). This research, then, coincides with previous studies which have found differences in teachers' use of feedback types in different subjects (e.g. Black et al. 2004; Nassaji & Wells 2000). Nonetheless, Hodgen and Marshall (2005), after comparing Mathematics and English classes (thought to be opposite) concluded that, in the end, what makes classrooms formative is the same regardless the subject: students being engaged, quality questioning, and extending students' thinking through justifications of their reasoning.

This chapter has also studied qualitatively student uptake following different feedback types. In line with previous studies (Lyster & Ranta 1997; Lyster 1998; Panova and Lyster 2002 – these studies refer to what I call *evaluation* as "recasts" and "explicit correction"; Wells 1993; Nassaji & Wells 2000), evaluative third turns limit students' opportunities for participation and taking the floor back, which is why they are said to move away from AfL pedagogy (Leung 2004). On the contrary, *re-route* is a kind of feedback that is almost always followed by student uptake (regardless of the subtypes), since the teacher is precisely waiting for this uptake to occur to see if, after the scaffolding provided in the *re-route*, students are able to get the correct answer. Again, these results are in the same line as those from previous research (Lyster & Ranta 1997; Panova & Lyster 2002). More specifically, uptake after *didactic recasts* which *focused on language* is modified output, which may influence L2 development and learning in a positive way (Lyster 2007). These findings are in line with results in other CLIL contexts but opposed to what happens in Immersion contexts (Llinares & Lyster 2014; Lyster & Mori 2006; Ellis & Sheen 2006; Oliver & Mackey 2003).

In addition, *expansion* has also been proven to be followed by student uptake quite often, many times in the form of long and complex contributions. When *expansion* took the form of a question, student uptake was very likely to take place. As shown above, these *expansions* in the form of questions are why-questions, that is, either *meta-cognitive questions* or *questions for reasons* which, as we saw in previous chapters, trigger quite complex student responses. As for *meta-feedback*, students' uptake is internal rather than external. That is, it is aimed at having an effect on students' mental and learning processes. Whether this internal uptake really occurs is not within the scope of this study, even though it would be very interesting for further research.

8.8 SUMMARY AND CONCLUSION

This last chapter on results has investigated teachers' feedback moves. It has been discovered that *evaluation* is the most frequent feedback move, in spite of the criticisms that using this move to evaluate has received in previous investigations. When we compare the two groups of teachers, it seems that AfL teachers really do align with AfL pedagogy, using more *expansion* and *meta-feedback* than Non-AfL teachers. As many researchers have claimed, this may have important consequences on students' learning, since with expansion students' contributions are elaborated, and with *meta-feedback*, their thinking is challenged and they are made aware of their improvements and weaknesses. Unexpectedly, there is no difference in the amount of *evaluation* used between the two groups of teachers. In addition, Non-AfL teachers, through their frequent use of *re-route*, also seem oriented to an AfL pedagogy at certain times. It has

also been unveiled that feedback types change depending on the subject (which is in line with previous research), and that even the same teacher teaching different subjects somehow changes their way of using the feedback move. Also very interesting has been the discovery of certain types of questions being followed more frequently by certain types of feedback: *questions for facts* followed by *evaluation*; *questions for reasons* by *expansion*; *meta-feedback* following *meta-questions*; and also some unexpected results, such as *questions for opinions* being followed by *evaluation* in most of the occasions (coinciding with results obtained by previous research). Feedback types in different subjects have been proved to be different, although there are also some commonalities, pointing to the subject as being an important variable. This was confirmed when types of feedback used by the same teacher in two different subjects were analysed: only in the case of Teacher 2 the variable subject seemed less powerful than the variable teacher itself. In other words, Teacher 2's AfL background seems to have a greater role than the subject specific pedagogies. Finally, a qualitative analysis of teacher's types of feedback and student uptake has shown that certain types of feedback prevent students from taking up the floor, such as *evaluation*, and others, on the contrary, encourage students to do so, as *re-route* and *expansion*.

9

General discussion and conclusion

Chapters 5-8 presented the results on each of interactional patterns that have been investigated in the present study (episodes, questions, responses, and feedback), including a discussion for each of these patterns. Those chapters also provided a comparative analysis within each of these patterns across AfL/Non-AfL classrooms, subjects, and individual teachers. In this final chapter, I present a general discussion which will revolve around the research questions and hypotheses posed in the introductory chapter. In light of these results, the chapter will move on to discussing the role of interaction in AfL CLIL classrooms, which will lead to a final proposal of an interactional model that could be used to develop AfL in these types of classes. The present chapter will also highlight pedagogical implications that can be applied to CLIL or AfL pedagogies. The chapter will conclude with an acknowledgement of the limitations of the study, and the formulation of ideas for follow-up studies that might be interesting and that might complete the present investigation.

9.1 INITIAL OBJECTIVES AND FINDINGS

This section will present the main findings in relation to the initial research questions and hypotheses of the investigation. The main goal was to discover how discourse shapes AfL in CLIL primary classrooms through various interactional patterns. The specific research questions were the following:

1. What are the interactional features and strategies which characterise AfL discourse in Primary CLIL classrooms?
2. Are these patterns specific for AfL classes or are they also found in similar classes where AfL is not implemented?

3. Are there differences in these interactional features and strategies that characterise AfL discourse across subjects? Are there differences in the discourse of the same teacher across different subjects?
4. To what extent do AfL strategies affect students' participation in classroom discourse?

The main hypotheses related to these questions were the following:

1. AfL is constructed through classroom discourse by the teacher and students through the use of certain types of episodes, questions and feedback.
2. There will be differences in the types of questions, types of episodes and types of feedback used by teachers in AfL schools and teachers in Non-AfL schools.
3. Differences will be found a) in the interactional patterns used in different subjects, and b) in the discourse of the same teacher teaching different subjects.
4. AfL strategies will positively influence students' participation and contributions. As a consequence, there will be differences in the types of responses given by students in AfL schools and Non-AfL schools.

First of all, in relation to the first research question, interaction has been analysed in primary school CLIL lessons taught by different teachers and in different subjects with the purpose of identifying features that make this interaction formative. The first of the features analysed was *teachers' questions*. As hypothesized (see hypothesis 1), the study has shown that there are certain types of questions which align with an AfL pedagogy and others that do not. *Questions for opinions*, *for reasons*, *meta-cognitive questions*, and *meta-questions* align with AfL pedagogy because they aim to explore students' thinking and learning processes, as shown in students' responses (see research question/hypothesis 1). In contrast, *questions for facts* do not align with such a pedagogy, since they focus on remembering facts rather than on extending students' reasoning and understanding (Black et al. 2003; Alexander 2004; Heritage 2007; Ruiz-Primo & Furtak 2007; Wragg & Brown 2001). The second feature analysed was *teachers' feedback*. The results showed that, in AfL lessons, *expansion*, *re-route*, or *meta-feedback* were frequent, in convergence which an AfL methodology, which encourages students to reflect on their learning or thinking processes. These types of feedback take students' responses into account and facilitate students' engagement, and hence, they are more aligned with AfL. In turn, *evaluation* often closes the exchange to give way to a new one, thus preventing students from uptake (Zhang Waring 2008; Wragg and Brown 2001; Black and Wiliam 1998a, b; Alexander 2004). On the other hand, feedback types such as *expansion*, *re-route*, or *meta-feedback* promote student uptake and/or push them to reflect about their learning and thinking processes (Ruiz-Primo & Furtak 2006; Wragg & Brown 2001; Black & Wiliam 1998a, b; Alexander

2004). In the same way, certain types of episodes are more characteristic of an AfL approach: *stating objectives for the lesson*, *explaining marks*, and *self-/peer-assessment*. These types of episodes seem to offer a good space for teachers to ask students *meta-questions* and offer them *meta-feedback*.

Secondly, in response to the second research question and hypothesis, there were significant differences between the classes in which AfL was implemented and the Non-AfL ones: in the former, teachers used significantly more of the question and feedback types that are more associated with a formative assessment pedagogy, that is, more *questions for opinions*, *for reasons*, *meta-cognitive*, and *meta-questions*, and more *expansion* and *meta-feedback*. As far as episodes are concerned, the ones that align with an AfL pedagogy (mentioned above) were almost exclusively found in AfL schools. However, some commonalities between types of schools were also observed: *whole-class discussion* and *classroom management and routines* were the most frequent episodes in both types; and *revision* and *explanation of activity/homework* are also present with similar percentages in both types of schools. We could argue, then, that those episodes characteristic of classroom pedagogy in general were more frequent, regardless of whether the school was implementing AfL or not.

As regards the third question and hypothesis, differences across subjects were also found. In the case of Science, both *questions for facts* and *for explanations* predominate, as well as *evaluation* and *expansion*, which indicates that content is frequently talked about and enquired as facts, and students are sometimes encouraged to elaborate and explain ideas. In Citizenship, the questions and feedback used by the teachers do not only evolve around facts but also aim at encouraging students to argue viewpoints, elaborate ideas, and reflect on learning processes. In Arts lessons, the results on question types and feedback types are somehow contradictory: whereas most questions are for *facts*, evaluation in the feedback move is not as frequent, and *meta-feedback*, *expansion*, and *re-route* are prominent. Regarding Drama lessons, students' viewpoints, personal opinions and reflections about assessment and learning are fostered through questions for *opinions*, *meta-cognitive questions*, and *meta-questions*. Likewise, apart from *evaluation*, feedback in Drama lessons also includes *expansion*, *re-route*, and *meta-feedback*, in line with the question types found. With these results in mind, it seems that Science, at least in the CLIL primary school context analysed, was the least favourable subject to AfL techniques. These results, of course, would need to be contrasted across educational and geographical levels. In a recent contrastive study on the use of appraisal and evaluative language by Finnish and Spanish CLIL students in secondary school biology classrooms, Llinares and Nikula (2016) observed differences in the participating roles of Spanish and Finnish students assigned by the teachers

and the effect of this on their language use, approach to content, and general participation in the classroom.

Also as part of the third research question and hypothesis, the analysis revealed that the fact that all teachers displayed differences regarding types of question and feedback in their two subjects supports the effect of the subject on the interaction patterns used. However, as not all teachers were involved in the same two subjects, we cannot categorically affirm that the differences are exclusively due to the specificities of the subject. The variable *teacher* could be very well playing its role too. In fact, it seems to be the case that both variables (*teacher* and *subject*) play a role. The only clear case in which types of feedback seem affected primarily by the subject is Arts, since hands-on activities led teachers to provide comments as to how well students were doing the work (*meta-feedback*), even though the two teachers teaching Arts were Non-AfL, and their overall percentage of *meta-feedback* was significantly lower than that of AfL teachers. AfL teachers were expected to align with an AfL discourse no matter which subject they were teaching. Contrary to expectation, both AfL and Non-AfL teachers display differences in their use of questions and feedback in different subjects. Consequently, it seems the case that the type of subject seems to play a bigger role than the variable *AfL* or *Non-AfL*. Episodes were also analysed in individual teachers and subjects. As it was already noted in the case of types of questions and feedback, in the case of episodes, it seems that both the variable *teaching style* and the variable *subject* play a role in the type of episodes used. This is more evident in the case of AfL teachers, since the episodes related to an AfL pedagogy are present in all their subjects.

In relation to the fourth research question and hypothesis of this study, the results show that question and feedback types which align with AfL methodology have a positive impact on students' participation and contributions. First, students' responses to questions which align with an AfL pedagogy were more complex, containing at least one *T-unit* which, in many cases, was made up of *more than one clause*, thus involving coordination or subordination. Likewise, results in this research indicate that it was more probable for students to have opportunities for uptake after teachers' feedback when this feedback is formative rather than evaluative.

9.2 THE ROLE OF INTERACTION IN AFL CLIL CLASSROOMS

The quality of interaction lies at the heart of good pedagogy (Black & Wiliam 1998a). This research has shown how different types of interactional patterns can lead to different levels of student participation and engagement, which could be expected to have different impacts on student learning. To provide further support in this regard, in this study we have shown that

some question and feedback types in IRF patterns can lead to a much more dialogic type of interaction, leading to longer and more complex students' turns. The present study has shown how certain types of questions and feedback in the IRFs encourage students' thinking (through the use of *meta-cognitive questions*, *meta-questions*, *questions for reasons*, *expansion*, and *re-route*), students pursuing their own ideas (by letting them initiate discourse and ask their own questions), and students reflecting on learning and assessment (through *meta-questions* and *meta-feedback*). This type of interaction was more frequent in CLIL classrooms in which AfL was implemented. Yet, as expected in any classroom context, IREs (Initiation-Response-Evaluation) also appeared in AfL classrooms. IREs are linked to authoritative approaches where just one voice and viewpoint is paid attention to, with no exploration of other ideas (Mortimer & Scott 2003), and associated with viewing teaching as a process of transmission (Barnes' Transmission Model, 1975), in which the teacher is a deliverer of content who passes information on to students. In Non-AfL classrooms, interaction was closer to recitation, which is still reported as the most common mode of teaching despite the benefits that responsive/contingent/dialogic teaching have for learning (Alexander 2004; Mortimer & Scott 2003). In recitation, *questions for facts* predominate and *questions for reasons*, *meta-cognitive*, and *meta-questions* are scarce. Far from discarding IREs and recitation patterns, the proposal in this study is for AfL types of interaction patterns to be integrated in any classroom, but particularly in CLIL, where opportunities for language use (for L2 language practice) and different ways of addressing content (for a better understanding of content taught in an L2) are perhaps particularly needed.

The present study has shown that whole-class discussions are the most frequent episodes in the data used for this research (along with class management episodes), both in AfL and Non-AfL classrooms. This interaction format is the rule in most classrooms (Lyster 2007: 87; Lyster & Mori 2006; Dalton-Puffer 2006; Fazio & Lyster 1998; Hiebert 1999; Alexander 2004), including CLIL classrooms (Nikula et al. 2013), and they are very important for developing an approach of inquiry in the classroom (Nassaji & Wells 2000). However, the study has also revealed that carrying out some types of episodes, such as *explaining marks* and *peer- and self-assessment*, can also be helpful not only when CLIL teachers specifically want to implement AfL but in general, as they offer a good context for *meta-questions* and *meta-feedback* to occur. In this way, students' learning and their awareness of that learning can offer crucial information for teachers, allowing them to adjust their teaching and decide on the next adequate teaching steps (Harrison & Howard 2009; Black & Wiliam 1998b). In addition, they provide excellent opportunities for "language through learning" (Coyle 2010), the use of evaluative language, and

an evaluative approach to content which are key for CLIL students content and language integrated learning (Morton & Llinares 2016; McCabe & Whittaker 2017)

With respect to the type of interaction found in different subjects, the present study has also unveiled that the type of subject itself plays a role when it comes to students' linguistic output. Specifically in AfL classes, students' output was more complex in Citizenship, since it is a subject that offers space for students to explore their own opinions and reflections, as opposed to Science, in which remembering facts and providing explanations to scientific phenomena is more frequent.

Finally, although learning has not been overtly treated in this study, the results have shown that participation in interaction provides opportunities for understanding and learning (Dalton-Puffer & Nikula 2006; Dalton-Puffer & Smit 2007; Dalton-Puffer 2009). It has been shown that AfL classrooms represent a more favourable context to promote learning, since students participate more and with longer turns, they often initiate discourse, they are asked to explain and argue their ideas and viewpoints, and they are encouraged to expand on their contributions. As Barnes (1975) argues, different types of communication lead to different types of learning. In this way, classrooms in which a lot of *questions for facts* are asked, trigger memorising and rote learning, as opposed to reasoning and deep learning, encouraged when teachers ask other types of questions (*questions for reasons, for explanations, meta-cognitive questions* etc.). This is related to Walsh's (2006; Seedhouse & Walsh 2010) concept of classroom interactional competence (CIC) and the identification of on-the-fly decisions made by teacher and students in interaction in order to create spaces for learning (adjusting interaction to meet the needs and goals). As shown in this study, creating these spaces for learning can be done through teachers' questions which aim at reasoning and deeper learning, letting students participate and contribute to discourse and offering *expansion*, *re-route*, and *meta-feedback* in the teachers' third move.

9.3 PROPOSAL FOR A MODEL OF AFL DISCOURSE PRACTICES IN CLIL CLASSROOMS

The following model is a proposal born from the present investigation based on the different analyses carried out on classroom interaction in AfL and Non-AfL contexts. This guide can be useful not only for the implementation of successful AfL practices in CLIL contexts, but also for increasing opportunities for more successful language and content interactional practices in CLIL classrooms in general.

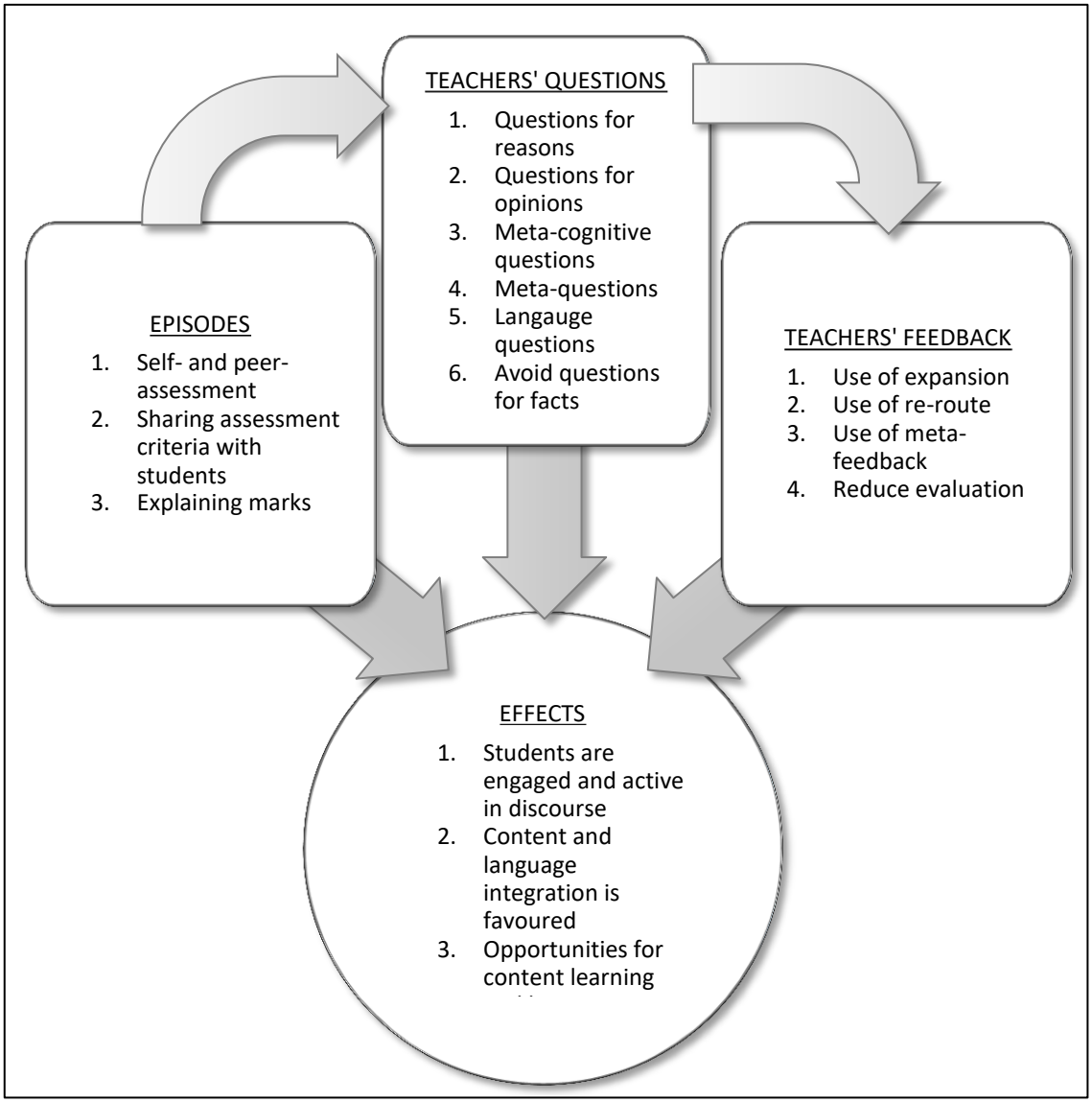


Figure 9.1 A model to implement interactive AfL in CLIL classrooms.

A successful implementation of AfL in CLIL requires teachers to make special use of certain types of questions: *questions for reasons*, *meta-cognitive questions*, and *meta-questions*. As seen in this study, with these question types, teachers can get the students to reason and think. In turn, *questions for facts*, which are necessary at school, need to be reduced, as students' engagement when these are asked has been proved to be lower. In addition to the above mentioned question types, teachers may also decide to use *language questions* when, in light of students' contributions, they feel some linguistic aspects need to be paid attention to. As far as feedback types are concerned, teachers should try to use *expansion*, *meta-feedback*, and *re-route* whenever students' interventions unveil misconceptions. Teachers can redirect students' thinking through the use of either *recasts* or *prompts*. The present results have shown that,

through the use of these feedback types, teachers encourage learners to reflect on their learning and assessment processes, as well as expand and build on students' contributions, which triggers student uptake and facilitates learning, both content learning and language learning. As proved in this investigation, through the use of the above mentioned question and feedback types, students are both active and engaged in the interaction and learning process. In the same way, if teachers use those kinds of questions and feedback, students' contributions, as shown by this study, are long and complex. This active participation in discourse also offers students opportunities to use the foreign language purposefully. As a result of the implementation of the proposed model, this research has shown that students can not only participate in discourse with long turns but also initiate discourse and perform a number of different functions when doing so. This not only makes the interaction more symmetric but it also enables students to find significant and relevant contexts that can aid their content and language learning. For this purpose, episodes such as *explaining marks*, *self- and peer-assessment*, and *sharing criteria with students* are especially relevant if students are to be active participants in their own learning and assessment processes. Therefore, they provide an appropriate context for teachers to ask *meta-questions* and provide *meta-feedback*, and for students to appropriate their learning.

9.4 PEDAGOGICAL APPLICATIONS

Different pedagogical implications can be drawn from the present study. One of the main implications is that the responsive nature of AfL lies in the quality of interaction, teacher questioning and feedback, and students' participation in discourse. These features are particularly relevant in CLIL classrooms for the integration of content and language, and for the double mediation that needs to exist in CLIL (Gibbons 2003). Also, very important in CLIL is to raise students' meta-cognitive processes, that is, to make them aware of how they learn to learn (Coyle 2006). This is closely linked to AfL, since students being active participants in their learning processes and being aware of them is a key feature in this approach (Black & Wiliam 1998a, b; Assessment Reform Group 2002). That is why *meta-questions* and *meta-feedback* would be very important in any CLIL classroom, whether or not AfL is being implemented. However, this does not happen so frequently when AfL is not present, as demonstrated by this research.

As shown by this investigation, AfL implementation significantly increases the quality of classroom interaction, a key component of any kind of learning, even more so in the case of CLIL, where content and language are expected to be learnt in an integrated fashion. That is why it is necessary to implement teacher training programmes in which the AfL interactional

approach explored in this thesis is introduced to CLIL teachers. As many authors have claimed, in AfL programmes, teachers need to renegotiate the learning contract (Perrenoud 1991; Black et al. 2003; Harrison & Howard 2009; Heritage 2010), which means that responsibility is passed on to students, and the teachers' aim is not only to transmit knowledge.

AfL teacher training programmes in countries such as the UK have proved to be successful but slow (Black et al. 2003; Black et al. 2004; Black and Wiliam 2003; Wiliam et al. 2004). For these programmes to be successful, it is paramount that the school community work together, that teachers and students have administrative support, and that all agents believe the investment and effort is worth it and will make a difference in students' future learning. Many times, the main barrier is the tensions and clashes existing between traditional and innovative teaching programmes, as well as the fact that formative assessment is often put in direct competition with summative assessment (Black & Wiliam 1998a; Rea-Dickins 2001). For the sake of education, nonetheless, reconciliations need to be found, and the two types of assessment should be complementary (Llinares et al. 2012; Black et al. 2011; Black 2012).

Teacher training programmes on AfL implementation can be beneficial for any classroom type (as already shown in other countries, like the UK – see Black & Wiliam 2003, Harrison & Howard 2009), but they would have an added value in the case of CLIL classrooms. Interaction is crucial to language learning; therefore, the higher the quality of interaction, the better opportunities for students to develop foreign language competence. At the same time, interaction in CLIL classrooms is also the means through which content learning is developed. As shown in this study, this interaction can be more engaging and meaningful if AfL is implemented. In other words, an AfL approach may allow the students' not only to learn content in a different way, but also to use the L2 in a wider range and more student-centred interactional practices. Lastly, since AfL is characterised by its responsiveness and contingency, it would help CLIL teachers to provide students with appropriate scaffolding at the two necessary levels, content and language, as well as to integrate both aspects in a meaningful way. As Llinares et al. (2012) put it, three characteristics of AfL are essential in CLIL: AfL is planned (therefore, not only knowledge and concepts but also language accompanying them need to be planned and included in the teaching, as well as genres and registers along with their grammar and vocabulary); AfL is reactive (instruction is adjusted); AfL is reciprocal (it fosters learner autonomy, and that can lead to L2 development).

Teacher training programmes in CLIL would not only have to include a focus on classroom interaction in general, but also an understanding of how classroom interaction works depending on the school subject. As this and other studies have shown, the subject area is an important

variable that needs to be taken into consideration (Black et al. 2004; Wiliam 2006; Black & Wiliam 1998a; Black & Wiliam 2009; Hodgen & Marshall 2005). This becomes especially relevant in CLIL, for pupils have to learn, in a foreign language, the different genres associated to the different subject areas (Dale & Tanner 2012; Llinares et al. 2012; Llinares & Whittaker 2009). As Llinares and Pascual (2014) point out, classroom discourse can help construct different genres, from complex scientific explanations in Science to developing personal opinions in Citizenship. Thus, interaction in CLIL classrooms is crucial if students are to become familiar with and to learn the different genres associated with different school subjects

9.5 LIMITATIONS OF THE STUDY

Every study has its limitations, since its scope needs to be narrowed down. One of the limitations one could argue about this research is that learning has not been treated as such. As Mortimer and Scott point out (2003: 101), "the analyses are carried out, and the findings reported, solely in terms of patterns of interaction, and the actual content of what is being taught and learned is not regarded as being a significant feature". In this study, students' participation and responses to different AfL techniques have been analysed, and conclusions have been raised regarding the opportunities of these contributions for content and language learning. Although students' participation is expected to enhance learning, the study has not proved that AfL techniques have actually improved students' content or language learning. As it was pointed out earlier in this study, learning is difficult to measure because, although it takes place in social interaction, it also involves cognitive processes to which no access is possible (Seedhouse 2010; Ellis 2010). Therefore, with classroom interaction, only inferences of learning can be made (Ellis 2010). However, more and more studies highlight the role of learning in social contexts and some of the instances shown in this study have illustrated learning moments. It is true that, in this study, learning could have been measured through pre- and post-tests (e.g. Ruiz-Primo & Furtak 2006). However, the dimensions of this study, including complete didactic units from different subjects taught by the same teachers in AfL and Non-AfL schools, would have made it difficult to create a pre- and post-test for each of the groups/contexts.

Another limitation of the study is that we don't have data of all the teachers teaching the same subjects and the same topics, which would have made comparability between them more reliable. However, getting access to schools is very difficult, even more when the objective is to record whole class sessions during complete didactic units twice a year. In addition, once access to the school is gained, not all teachers are ready to or can collaborate in the project. Previous experience and professional knowledge of teachers was another aspect that could not be

controlled and that it may have had some influence on the results. Three of the teachers were pretty close in age and experience, but the other one had far more teaching experience than the rest. In spite of these problems, the data collected is very rich, since they consist of complete didactic units and include different teachers and subjects, which has allowed for very detailed and varied analyses.

Although the study has identified teachers' and students' interactional practices across episodes, it would have been interesting to have explored in more detail the development of AfL techniques throughout the lessons from the beginning to the end of each didactic unit. Even though this can be done in future studies, the length of the units recorded was so variable (from 1 lesson up to 5 or 6) that results coming out of that type of analysis may not have been reliable. In the same way, if the topics of the didactic units in each subject had been similar, it would have helped this type of analysis be more solid. Nonetheless, having complete didactic units has allowed us to have extensive and reliable data for all the types of discourse analyses carried out in this research.

It would have also been interesting to include a detailed analysis of group talk, especially in CLIL classes in which AfL was implemented, in which these episodes happen more frequently. As explained in Chapter 4, these episodes were only further analysed in terms of IRF patterns whenever there was interaction with the teacher about content, language, or both. However, it is true that a more detailed analysis of peer interaction would be needed in order to complete the whole analysis of classroom interaction. This type of talk should not be underestimated or ignored. On the contrary, its importance should be emphasized (Barnes 1975).

Finally, interviews with teachers after each recorded unit were conducted, as they are a valuable exercise of reflection for teachers about their practice. They were not used in this dissertation because the focus was on classroom interaction. Nonetheless, such interviews can be analysed for further research, in order to give an idea about to what extent teachers practice AfL strategies or strategies that move away from this pedagogy consciously.

9.6 FURTHER RESEARCH

With this research as a starting point, there are many other ideas to be considered for further research. For instance, the number of subjects could be extended. In this research, we have dealt with four different school subjects, but these can be increased. This could lead to the construction of subject-specific AfL pedagogies, that is, how AfL can help develop the different

genres and cognitive discourse functions (Dalton-Puffer 2013, 2016) required in different subjects in CLIL classes.

It could also be interesting to extend the data and unveil differences (if any) and similarities in the implementation of AfL in CLIL classes, in L1 content classes, and in EFL classrooms. In this way, it could be discovered whether specific characteristics of different types of contexts (L1 content/L2 content or EFL) may have an effect on AfL interaction and its implementation, with possible contributions for the design of context-specific AfL pedagogies.

Apart from extending the data to more subjects and more types of contexts, data from more academic levels could also be compiled. The present study has focused on upper primary (fifth and sixth year). A similar study but in CLIL secondary education would also be very interesting, to see whether the implementation of AfL would be different in terms of classroom discourse at this academic level, whether the co-construction of AfL has the same features as in primary classrooms, how the implementation of AfL in CLIL Secondary Education affects students' learning, etc. In Spain, however, it is hard to find secondary schools that implement AfL. Consequently, such a study would be very difficult to carry out.

In addition, a detailed quantitative analysis of IRF patterns within the different types of episodes highlighted in this research would provide a bigger picture of classroom interaction, contributing to making sound conclusions on whether different episodes trigger certain types of teachers' questions and feedback, and consequently, affect student engagement in such interaction.

With the AfL implementation model proposed in this research, along with results from other current and future studies, teacher training programmes to implement AfL pedagogy could be launched. After some time of implementation (1-2 years), a research study can be led to find out what positive effects it has had on both teachers and students (see Harrison & Swaffield 2003 on the KMOFAP Project).

As pointed out above, in the present study, learning as such has not been the main focus. Learning can only be inferred from interaction or one can only claim that interaction provides opportunities for learning. For further investigation, it would be interesting to add tests (pre- and post-) in order to be able to measure in a more objective way what students have or have not learnt (Ruiz-Primo & Furtak 2006). However, this was not the purpose of the present study. Students' learning of content and/or language in a Non-AfL class could be compared to students' learning in an AfL class, and see if differences are statistically significant. Should significant differences arise in favour of the AfL group, one of the possible causes may be the implementation of AfL.

As was mentioned in the previous section, although not used, teacher interviews were conducted. Future analysis of these interviews could add a new dimension to the study. As Cowie and Bell (1999) found out, in their study teachers implemented AfL but were not aware of it. Interviews with teachers could be a useful way to encourage more reflection on their part, enabling them to become aware of what they do (Cowie & Bell 1999). In addition, information could be obtained not only regarding AfL teachers' awareness of their interactional AfL techniques but also regarding Non-AfL teachers' interactional practices, which sometimes revealed AfL features as well.

Similarly, conducting interviews with students about their impressions on their learning and motivation would be very important. At the end of the day, AfL is mainly about students and their learning, and so their views on an AfL pedagogy should be as valuable as teachers'. Ultimately, for further research on AfL, involving students and listening to them should be crucial.

9.7 CONCLUDING REMARKS

In light of this research, it was concluded that implementing AfL in CLIL classrooms can have important benefits for the type of interaction that goes on inside the black box, as this interaction combines the learning of content through a foreign language and the learning of using that foreign language appropriately. The resulting interaction in AfL CLIL classes is more dialogic and responsive, more aligned with teaching as a process of enquiry rather than with teaching as a process of transmission. This type of interaction is also better for the integration of content and language and with assessment, as some researchers have already emphasized (Llinares et al. 2012).

The results presented throughout this dissertation have led me to propose a model to implement AfL in classroom interaction in CLIL classrooms. This model consists of teachers' questions to make students think and reason; teachers' feedback to build and expand on students' contributions and make them reflect about learning; and episodes which trigger reflection about the learning process.

This research has intended to fill various gaps in CLIL and AfL investigation. From the AfL perspective, there are not so many studies focusing on AfL and how it is developed in classrooms, in spite of its importance (Black & Wiliam 1998a; Leung 2004; Leung & Davison 2009). More specifically in CLIL, AfL has barely been researched (but see Llinares et al. 2012 for a brief introduction). An exception is Basse (2016), who investigated AfL in CLIL

classrooms in relation to teacher motivational L2 strategies and student motivation and meta-cognitive abilities using the same corpus as the one used in the present study. Her results showed that AfL teachers used L2 motivational strategies more frequently and in a more varied way, which resulted in a more motivational discourse. As for students' motivation, there were no significant differences between AfL and Non-AfL schools, although lower achieving students in AfL schools reflected more critically than those in Non-AfL schools when assessing themselves. The present research is, then, one of the first studies to focus on AfL practices in CLIL contexts, and the first to investigate AfL CLIL classroom interaction, focusing on the interaction going on inside the black box (Dalton-Puffer & Smit 2007). From the point of view of CLIL, it is indeed very important for CLIL to have studies addressing the actual teaching and learning in the classrooms, as claimed recently by Cenoz, Genesee and Gorter (2014). Finally, the current investigation contributes to fill the gap existing in CLIL studies at the primary level, and more specifically, interaction in primary CLIL classes.

As a conclusion, the present study has been a contribution to both CLIL and AfL research and practice, adopting a discourse perspective. Through an in-depth analysis of classroom interaction jointly constructed by teacher and students, it has highlighted the main characteristics of AfL discourse in CLIL classes as opposed to Non-AfL discourse in these classes. The types of questions and feedback used by teachers affect the types of responses given by students as well as student uptake. Using certain types of questions and feedback and not others helps construct a responsive, dynamic, and contingent type of interaction. This investigation has also addressed the integration of content and language (and element that makes CLIL classrooms different), and how the implementation of AfL can be helpful in this matter. Finally, this research has contributed to demonstrate how different subjects may require different types of discourse or genres.

References

Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*, 12(1), 75-93.

Aikenhead, G. (1997). A framework for reflecting on assessment and evaluation. In *Globalization of science education: International conference on science education* (pp. 195 199). Seoul, Korea: Korean Educational Development Institute.

Airasian, P. W. (1997). *Classroom assessment*. McGraw Hill: New York. 3rd edition.

Airey, J. (2009). Estimating undergraduate bilingual scientific Literacy in Sweden. *International CLIL Research Journal*, 1(2), 26-35.

Alexander, R. J. (2004). *Towards dialogic teaching: rethinking classroom talk*. Cambridge: Dialogos.

Aljaafreh, A., & Lantolf, J. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal*, 78, 465-483.

Allen, P., Swain, M., Harley, B., & Cummins, J. (1990). Aspects of classroom treatment: Toward a more comprehensive view of second language education. In B. Harley, P. Allen, J. Cummins, & M.Swain (Eds.), *The development of bilingual proficiency* (pp. 57-81). Cambridge: Cambridge University Press.

Allwright, D. (1975). Problems in the study of the language teacher's treatment of learner error. In M. Burt, & H. Dulay (Eds.), *New directions in second language learning, teaching and bilingual education: On TESOL '75* (pp. 96-109). Washington: TESOL.

Allwright, D., & Bailey, K. (1991). *Focus on the Language Classroom: An Introduction to Classroom Research for Language Teachers.* Cambridge: Cambridge University Press.

Ames, C. (1992). Classrooms: goals, structures and student motivation. *Journal of Educational Psychology*, 84, 261-271.

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, 28, 543-574.

Anderson, L. W., Kraftwohl, D. R., et al. (Eds.) (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Anderson, K. T., Zuiker, S. J., Taasoobshirazi, G., & Hickey, D. T. (2007). Classroom discourse as a tool to enhance formative assessment and practise in Science. *International Journal of Science Education*, 29(14), 1721-1744.

Andrews, J. (2004). Book Review: Investigating formative assessment: teaching, learning and assessment in the classroom. *Language Testing*, 21(3), 432-435.

Anghel, A., Cabrales, A., & Carro, J. M. (2016). Evaluating a bilingual education program in Spain: the impact beyond foreign language learning. *Economic Inquiry*, 54(2), 1202-1223.

Appel, J. (2010). Participation and instructed language learning. In P. Seedhouse, S. Walsh, & C. Jenks (Eds.), *Conceptualising learning in applied linguistics* (pp. 206-224). Basingstoke: Palgrave Macmillan.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-Based Approaches to Developing Understanding: Classroom Instruction and Student Performance in Middle and High School English. *American Educational Research Journal*, 40(3), 685-730.

Arkoudis, S., & O'Loughlin, K. (2004). Tensions between validity and outcomes: Teachers' assessment of written work of recently arrived immigrant ESL students. *Language Testing*, 20, 284-304.

Assessment Reform Group (1999). *Assessment for Learning: beyond the black box*. Cambridge: University of Cambridge School of Education. Retrieved on January 20, 2012 from http://arg.educ.cam.ac.uk/AssessInsides.pdf.

Assessment Reform Group (2002). *Assessment for learning: 10 principles*. Retrieved January 20, 2012 from http://www.assessment-reformgroup.org/publications.html.

Aston, G. (1986). Troubleshooting in interaction with learners: The more the merrier? *Applied Linguistics*, 7, 128-143.

Aulls, M. W. (1998). Contributions of classroom discourse to what content students learn during curriculum enactment. *Journal of Educational Psychology*, 90(1), 56-69.

Badertscher, H., & Bieri, T. (2009). *Wissenserwerb im content and language integrated learning: Empirische Befunde und Interpretationen* [Knowledge acquisition in content-and-language- integrated learning: Empirical evidence and interpretations]. Bern/Stuttgart/Wien: Haupt.

Balzer, W. K., Doherty, M. E., & O'Connor, R., Jr. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, 106(3), 410-433.

Barnes, D. (1969). Language in the secondary classroom. In D. Barnes, J. Britton, & M. Torbe (Eds.), *Language, the Learner and the School* (pp. 11-77). Harmondsworth: Penguin Books.

Barnes, D. (1975). *From communication to curriculum*. London: Penguin books.

Barnes, D. (1976). *From Curriculum to Communication*. Harmondsworth: Penguin.

Barnes, B. (1982). On the extension of concepts and the growth of knowledge. *Sociological Review*, 30, 303-333.

Barnes, D. (1992). *From Communication to Curriculum*. Portsmouth, NH: Boynton/Cook.

Basse, R. (2016). Assessment for learning in primary bilingual classrooms in Spain: a corpus-based linguistic analysis of teacher motivational strategies and student metacognitive abilities. Unpublished doctoral dissertation. Madrid: Universidad Autónoma de Madrid.

Batstone, R. (Ed.). (2010). *Sociocognitive perspectives on language use and language learning*. Oxford: Oxford University Press.

Bell, B., & Cowie, B. (1999) A model of formative assessment in science education. *Assessment in Education*, 6(1), 101-116.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85: 536-553.

Berg, B. L. (2009). *Qualitative research methods for the social sciences*. 7th ed. Boston, MA: Allyn & Bacon.

Black, P. J. (1993). Formative and summative assessment by teachers. *Studies in Science Education*, 21, 49-97.

Black, P. J. (2001). Formative assessment and curriculum consequences. In D. Scott (Ed.), *Curriculum and assessment* (pp. 7-23). Westport, CT: Ablex Publishing.

Black, P. J. (2003). The nature and value of formative assessment for learning . Paper presented at AERA Chicago, 22 April, Presidential invited session 34.011. http://www.kcl.ac.uk//depsta/education/hpages/pblackpubs.html.

Black, P. J. (2012). Formative and summative aspects of assessment: theoretical and research foundations in the context of pedagogy. Draft for Sage/McMillan chapter.

Black. P. J., & Harrison, C. (2001a). Feedback in questioning and marking: The science teacher's role in formative assessment. *School Science Review*, 83(301), 55-61.

Black, P. J., & Harrison, C. (2001b). Self-Assessment, Peer-Assesment and Taking Responsibility: the Science Student's Role in Formative Assessment. *School Science Review*, 83(302), 43-49.

Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Maidenhead: Open University Press.

Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 9-21.

Black, P. J., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), 451-469.

Black, P. J., & Wiliam, D. (1998a). Assessment and Classroom Learning. *Assessment in Education*, 5(1), 7-74.

Black, P. J., & Wiliam, D. (1998b). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.

Black, P. J. & Wiliam, D. (2003). In praise of educational research': formative assessment. *British Educational Research Journal*, 29, 623-38.

Black, P. & Wiliam, D. (2006) Developing a theory of formative assessment, In J. Gardner (Ed.), *Assessment and learning*. London, Sage.

Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessessment, Evaluation and Accountability*, 21, 5-31.

Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Taylor (Ed.), *Educational evaluation: New roles, new means: The 68th yearbook of the National Society for the Study of Evaluation, Part II* (pp. 26-50). Chicago: University of Chicago Press.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain.* New York: David McKay.

Bloome, D., Beierle, M., Grigorenko, M, & Goldman, S. (2009). Learning over time: Uses of intercontextuality, collective memories, and classroom chronotopes in the construction of learning opportunities in a ninth-grade language arts classroom. *Language and Education*, 23(4), 313-34.

Blote, A. W. (1995). Students' self-concept in relation to perceived differential teacher treatment. *Learning & Instruction*, 5(3), 221-236.

Bond, L., Smith, R., Baker, W. K., & Hattie, J. A. (2000). *Certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*. Washington, DC: National Board for Professional Teaching Standards.

Bonnet, A. (2004). *Chemie im bilingualen Unterricht — Kompetenzerwerb durch Interaktion*. [Chemistry in bilingual content teaching — Skills acquisition through interaction]. Opladen, Germany: Leske und Budrich.

Boyd, M., & Maloof, V. M. (2000). How teachers can build on student-proposed intertextual links to facilitate student talk in the ESL classroom. In J. K. Hall & L. S. Verplaetse, *Second and foreign language learning through classroom interaction* (pp. 139-155). Mahwah, NJ: Lawrence Erlbaum Associates.

Breen, M., Barratt-Pugh, C., Derewianka, B., House, H., Hudson, C., Lumley, T., et al. (1997). *Profiling ESL children: How teachers interpret and use national and state assessment frameworks*. Canberra, Australia: Department of Employment, Education, Training and Youth Affairs.

Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: a review of issues. *Language Testing*, 15, 45-85.

Brindley, G. (2001). Outcomes-based assessment practice: Some examples and emerging insights. *Language Testing*, 18, 393-407.

Broadfoot, P. M. (1996). *Education, assessment and society: a sociological analysis*. Buckingham: Open University Press.

Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education*, 11(1), 7-27.

Broadfoot, P., Pollard, A., Osborn, M., McNess, E., & Triggs, P. (1998). Categories, standards and instrumentalism: theorizing the changing discourse of assessment policy in English primary education. Paper presented at the Annual Meeting of the American Educational Research Association, 13-17 April, San Diego, California, USA.

Brock, C., Crookes, G., Day, R., & Long, M. (1986). The differential effects of corrective feedback in native speaker/non-native speaker conversation. In R. R. Day (Ed.), *"Talking to learn": Conversation in second language acquisition* (pp. 327-351). Rowley, MA: Newbury House.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12.

Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106, 429-458.

Brookhart, S. M. (2009). Editorial: Special issue on the validity of formative and interim assessment. *Educational Measurement: Issues and Practice,* 28(3), 1-4.

Brooks, F. B., Donato, R., & McGlone, J. V. (1997).When are they going to say "it" right? Understanding learner talk during pair-work activity. *Foreign Language Annals*, 30(4), 524-541.

Brophy, J. E., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328-375). New York: Macmillan.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology*, 79, 474-482.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1-14.

Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, 78(3), 210-216.

Byram, M. (1989). *Cultural studies in foreign language education*. Philadelphia: Multilingual Matters.

Byrnes, H. (2008). Assessing content and language. In E. Shohamy, & N. H. Hornberger (Eds), *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment 2nd ed*. (pp. 37-52). Boston, MA: Springer.

Cameron, L., Moon, J., & Bygate, M. (1996). Language development of bilingual pupils in the mainstream: How do pupils and teachers use language? *Language and Education*, 10, 221-36.

Canale M., & Swain M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.

Carless, D. (2007). Conceptualizing pre-emptive formative assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 171-184.

Carlsen, W. S. (1991). Questioning in classrooms: A sociolinguistic perspective. *Review of Educational Research,* 61, 157-178.

Cazden, C. B. (1986). Classroom discourse. In Wittrock, M. C. (Ed.), *Handbook of Research on Teaching* (pp. 432-463). New York: MacMillan Publishing Company.

Cazden, C. B. (1988). *Classroom discourse. The language of teaching and learning*. Portsmouth, NH: Heinemann.

Cazden, C. B. (2001). *Classroom discourse. The language of teaching and learning* (2nd ed.). Portsmouth, NH: Heinemann.

Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, 35(3), 243-262.

Chamot, A. U., & O'Malley, J. M. (1987). The Cognitive Academic Language Learning Approach: A Bridge to the Mainstream. *TESOL Quarterly*, 21(2), 227-249.

Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning*, 27, 29-46.

Chaudron, C. (1985). Comprehension, comprehensibility and learning in the second language classroom. *Studies in Second Language Acquisition*, 7, 216-232.

Chaudron, C. (1986). Teachers' priorities in correcting learners' errors in French immersion classes. In R. Day (Ed.), *Talking to learn* (pp. 64-84). Rowley, MA: Newbury House.

Christie, F. (2002). *Classroom discourse analysis: A functional perspective*. London, UK: Continuum.

Christie, F., & Derewianka, B. (2008). *School discourse: Learning to write across the years of schooling*. London: Continuum.

Clarke, D. J. (1988). Realistic assessment. In D. Firth (Ed.), *Maths counts—who cares?* (pp. 187-192). Parkville, Vic.: Mathematical Association of Victoria.

Clarke, D. J. (1998). Studying the classroom negotiation of meaning: Complementary accounts methodology. In A. Teppo (Ed.), *Qualitative research methods in mathematics education*

(pp. 98-111). [Monograph Number 9 of the *Journal for Research in Mathematics Education*.] Reston, VA: NCTM.

Coffin, C. (2006). *Historical discourse.* London: Continuum.

Consolo, D. (2000). Teachers' action and student oral participation in classroom interaction. In J. K. Hall, & L. S. Verplaetse (Eds.), *Second and foreign language learning through classroom interaction* (pp. 91-108). Mahwah, NJ: Lawrence Erlbaum Associates.

Cortés-Conde, D. (2000). Identity and Ideology: Culture and pragmatics in content-based ESL. In J. K. Hall, & L. S. Verplaetse (Eds.), *Second and foreign language learning through classroom interaction* (pp. 173-187). Mahwah, NJ: Lawrence Erlbaum Associates.

Cowan, J. (2006). *On becoming an innovative University teacher: reflection in action.* 2nd edn. Oxford: Oxford University Press.

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education*, 6(1), 102-116.

Coyle, D. (2006). Motivating teachers motivating learners through content and language integrated learning. *Scottish Modern Languages Review*, 13, 1-18.

Coyle, D. (2007). Content and Language Integrated Learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism*, 10(5), 543-62.

Coyle, D., Hood, P., & Marsh, D. (2010). *Content and Language Integrated Learning.* Cambridge: Cambridge University Press.

Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *Journal of Educational Psychology*, 83, 17-27.

Creese, A. (2005). Is this content-based language teaching? *Linguistics and Education*, 16(2), 188-204.

Crooks, T. J. (1988). The impact of classroom evaluation on students. *Review of Educational Research*, 5, 438-481.

Crooks, T. J. (2006). Assessment for learning in New Zealand: What's happening and what's next? Paper presented at the What's next with assessment for learning conference. Massey University, Palmerston North, New Zealand.

Crystal, D. (1991). *A dictionary of Linguistics and Phonetics*. Oxford: Basil Blackwell.

Cummins, J. (1979). Linguistic interdependence and the education of bilingual children. *Review of Educational Research*, 49(2), 221-251.

Cummins, J. (1984). *Bilingualism and special education: issues in assessment and pedagogy*. Clevedon, England: Multilingual Matters.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In *Schooling and language minority students: A theoretical framework* (pp. 3-49). Los Angeles: California State University, National Evaluation, Dissemination and Assessment Center.

Cummins, J. (1992). Language proficiency, bilingualism, and academic achievement. In P. A. Richard-Amato, & M. A. Snow (Eds.), *The multicultural classroom: Readings for content-area teachers* (pp. 16-26). New York: Longman.

Cummins, J. (2004). Language and literacy in bilingual children. *Journal of Child Language*, 31(2), 424-429.

Dafouz Milne, E., & Núñez Perucha, B. (2010). Metadiscursive devices in university lectures. A contrastive analysis of L1 and L2 teacher performance. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 213-231). Amsterdam, The Netherlands: John Benjamins.

Dale, L., & Tanner, R. (2012). *CLIL activities. A resource for subject and language teachers*. Cambridge: CUP.

Dalton-Puffer, C. (2005). Negotiating interpersonal meanings in naturalistic classroom discourse: Directives in Content and Language Integrated Classrooms. *Journal of Pragmatics*, 37(8), 1275-93.

Dalton-Puffer, C. (2006). Questions in CLIL classroom: Strategic questioning to encourage speaking. In A. Martínez-Flor, & E. Usó (Eds.), *Current trends in the development of the four skills within a communicative framework*. Berlin: Mouton de Gruyter, 187-213.

Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated Learning (CLIL) Classrooms*. Amsterdam: John Benjamins.

Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): Current research from Europe. In W. Delanoy, & L. Volkmann (Eds.), *Future perspectives for English language teaching* (pp. 139-157). Heidelberg, Germany: Carl Winter.

Dalton-Puffer, C. (2009). Communicative competence and the CLIL lesson. In Ruiz de Zarobe, & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 197-214). Bristol: Multilingual Matters.

Dalton-Puffer, C. (2011). Content-and-Language Integrated Learning: from practice to principles? *Annual Review of Applied Linguistics* 31, 182-204.

Dalton-Puffer, C. (2013). A Construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics* 1(2), 1-38.

Dalton-Puffer, C. (2016). Cognitive Discourse Functions: Specifying an Integrative Interdisciplinary Construct In Nikula, T., Dafouz, E. Moore, P. & U. Smit (Eds.), *Conceptualising integration CLIL and multilingual education* (pp. 29-54). Clevedon: Multilingual Matters.

Dalton-Puffer, C., Huttner, J., Jexenflicker, S., Schindelegger, V., & Smit, U. (2008). *Content and language integrated learning an Osterreichs Höheren Technischen Lehranstalten.* Forschungsbericht. Vienna, Austria: Universitat Wien & Bundesministerium für Unterricht, Kultur und Kunst.

Dalton-Puffer, C., Llinares, A., Lorenzo, F., & Nikula, T. (2014). "You Can Stand Under My Umbrella": Immersion, CLIL and Bilingual Education. A Response to Cenoz, Genesee & Gorter (2013). *Applied Linguistics*, 35(2), 213-218.

Dalton-Puffer, C., & Nikula, T. (2006). Pragmatics of content-based instruction: Teacher and student directives in Finnish and Austrian classrooms. *Applied Linguistics*, 27, 241-267.

Dalton-Puffer, C., Nikula, T., & Smit, U. (Eds.) (2010). *Language use and language learning in CLIL classrooms*. Amsterdam: John Benjamins.

Dalton-Puffer, C., & Smit, U. (Eds.) (2007). *Empirical perspectives on CLIL classroom discourse*. Frankfurt: Peter Lang.

Dassa, C. (1990). From a horizontal to a vertical method of integrating educational diagnosis with classroom assessment. *Alberta Journal of Educational Research,* 36, 35-44.

Davis, B. (1997). Listening for differences: An evolving conception of mathematics teaching. *Journal for Research in Mathematics Education,* 28(3), 355-376.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393-415.

Davison, C., & Williams, A. (Eds.). (2002). *Learning from each other: Literacy, labels and limitations. Studies of child English language and literacy development K-12, Volume 2.* Melbourne: Language Australia.

Day, E. M., & Shapson, S. M. (1996). *Studies in immersion education*. Clevedon: Multilingual Matters.

De Graaff, R., Koopman, G. J., Anikina, Y., & Westhoff, G. (2007). An observation tool for effective L2 pedagogy in Content and Language Integrated Learning (CLIL). *International Journal of Bilingual Education and Bilingualism*, 10(5), 603-624.

Dempster, F. N. (1992). Using tests to promote learning: a neglected classroom resource. *Journal of Research and Development in Education*, 25, 213-17.

Deutsch, M. (1979). Education and distributive justice. *American Psychologists*, 34(5), 391-401.

Dillon, J. T. (1988). *Questioning and discussion: a multi-disciplinary approach.* New York: Ablex.

Dillon, J. T. (1990). *The practice of questioning*. London, Routledge.

Dinsmore, D. (1985). Waiting for Godot in the EFL classroom. *ELT Journal*, 39(4), 225-234.

Dohrn, E., & Bryan, T. (1994). Attribution instruction. *Teaching Exceptional Children*, 26(4), 61-63.

Donato, R. (1994). Collective scaffolding in second language learning. In J. Lantolf, & G. Appel (Eds.) *Vygotskian approaches to second language research* (pp. 33-56). Norwood, NJ: Ablex.

Donato, R. (2000). Sociocultural contributions to understanding the foreign and second language classroom. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 27-50). New York: Oxford University Press.

Doughty, C. (1992). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies In Second language Acquisition*, 13, 431-70.

Doughty, C. (1994). Finetuning of feedback by competent speakers to language learners. In J. Alatis (Ed.), *Strategic interaction and language acquisition: Theory, practice, and research. GURT 1993* (pp. 96-108). Washington: Georgetown University Press.

Doughty, C., & Varela, E. (1998) Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom. Second language acquisition* (pp. 114-138). Cambridge: Cambridge University Press.

Doughty, C., & Williams, J. (Eds.) (1998). *Focus on Form in Classroom. Second Language Acquisition.* Cambridge: Cambridge University Press.

Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (1992). *Discourse transcription*. Santa Barbara Papers in Linguistics 4.

Duff, P.A. (2000). Repetition in foreign language classroom interaction. In J. K. Hall, & L. S. Verplaetse (Eds.), *Second and foreign language learning through classroom interaction* (pp. 92-117). Mahwah, NJ: Lawrence Erlbaum Associates.

Duschl, R. A. (2003). Assessment of inquiry. In J. M. Atkin, & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41-59). Arlington, VA: National Science Teachers Association Press.

Duschl, R. A., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment*, 4(1), 37-73.

Dweck, C. (2000). Self theories: their role in motivation, personality, and development. Philadelphia, PA: Psychology Press.

Dwyer, C. A. (1998). Assessment and classroom learning: Theory and practice. *Assessment in Education*, 5(1), 131-137.

Earley, P. C., Northcraft, G. B., Lee, C., & Lituchy, T. R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal*, 33(1), 87-105.

Easley, J. A., & Zwoyer, R. (1975). Teaching by listening. *Contemporary Education*, 57(1),19-25.

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: the concept of "diagnostic competence". *Language Testing*, 21(3), 259-283.

Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London: Routledge.

Edwards, A.D., & Westgate, D. P. G. (1994). *Investigating classroom talk* (2nd ed.). London: Westgate.

Ellis, R. (1984). *Classroom second language development*. Oxford: Pergamon Press.

Ellis, R. (1994). *The Study of Second Language Acquisition.* Oxford: Oxford University Press.

Ellis, R. (1997). *SLA research and language teaching.* Oxford: Oxford University Press.

Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24, 143-188.

Ellis, R. (2010a). Cognitive, social, and psychological dimensions of corrective feedback. In R. Batstone (Ed.), *Sociocognitive perspectives on language use and language learning* (pp. 151-165). Oxford: Oxford University Press.

Ellis, R. (2010b). Theoretical pluralism in SLA: Is there a way forward? In P. Seedhouse, S. Walsh, and C. Jenks (Eds.), *Conceptualising learning in applied linguistics* (pp. 23-51). Basingstoke: Palgrave Macmillan.

Ellis, R., Basturkmen, H., & Loewen, S. (2001a). Learner uptake in communicative ESL lessons. *Language Learning*, 51, 281-318.

Ellis, R., Basturkmen, H., & Loewen, S. (2001b). Pre-emptive focus on form in the ESL class-room. *TESOL Quarterly*, 35, 407-32.

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21, 285-301.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28, 339-368.

Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, 28(4), 575-600.

Ellis, R., & Sheen, Y. (2011). Corrective feedback in language teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Volume II, pp. 593-610). New York: Routledge.

Ellis, R., Tanaka, T., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning*, 44, 449-91.

Escobar Urmeneta, C. (2007). Mesurament de competències lingüístiques en aules AICLE, en contextos multilingües: allò que els números amaguen. In O. Guasch, & L. Nussbaum (Eds.), *Aproximacions a la noció de competència multilingüe* (pp. 133-146). Bellaterra: Servei de Publicacions de la UAB.

Escobar Urmeneta, C., & Pérez-Vidal, C. (2004). Teacher education for the implementation of a Content and Language Integrating Learning approach (CLIL) in the school system. In R. Wilkinson (Ed.), *Integrating Content and Language. meeting the challenge of multilingual education* (pp. 402-415). Maastricht: Universitaire Pers Maastricht.

Eurydice (2006). *Content and language integrated learning (CLIL) at school in Europe.* Brussels: European Commission. Retrieved from http://eacea.ec.europa.eu/education/eurydice/ thematic_studies_archives_en.php#2006 (April 2011).

Evnitskaya, N. (2012). *Talking science in a second language: The interactional co-construction of dialogic explanations in the CLIL science classroom.* Unpublished Doctoral Dissertation. Universitat Autònoma de Barcelona.

Evnitskaya, N., & Morton, T. (2011). Knowledge construction, meaning-making and interaction in CLIL science classroom communities of practice. *Language and Education*, 25(2), 109-127.

Fanselow, J. F. (1977). Beyond Rashomon-conceptualizing and describing the teaching act. *TESOL Quarterly* 11(1), 17-39.

Fazio, L., & Lyster, R. (1998). Immersion and submersion classrooms: A comparison of instructional practices in language arts. *Journal of Multilingual and Multicultural Development*, 19, 303-317.

Fernández Fontecha, A. (2009). Spanish CLIL: research and official actions. In Y. Ruiz de Zarobe, & R. M. Jiménez Catalán (Eds.), *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 3-21). Bristol: Multilingual Matters.

Fontana, D., & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology*, 64, 407-417.

Foos, P. W., Mora, J.J., & Tkacz, S. (1994). Student Study Techniques and the Generation Effect. *Journal of Educational Psychology*, 86, 567-576.

Foster, P., & Ohta, A. (2005). Negotiation of meaning and peer assistance in second language classrooms. *Applied Linguistics*, 26, 402-430.

Frigols Martin, M. J., Marsh, D., & Naysmith, J. (2007). Competence-building for teachers of CLIL: vocational education. In D. Marsh, & D. Wolff (Eds.), *Diverse contexts-converging goals. CLIL in Europe* (pp. 33-45). Frankfurt: Peter Lang.

Frölich, M., Spada, N., & Allen, P. (1985). Differences in the communicative orientations of L2 classrooms. *TESOL Quarterly*, 19(1), 27-57.

Fontana, D., & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology*, 64, 407-417.

Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making Students' Thinking Explicit in Writing and Discussion: An analysis of formative assessment prompts. *Science Education*, 92(5), 798-824.

Gajo, L. (2007). Linguistic knowledge and subject knowledge: How does bilingualism contribute to subject development? *The International Journal of Bilingual Education and Bilingualism*, 10(5), 563-579.

Gallardo del Puerto, F., Gómez Lacabex, E., & García Lecumberri, M. L. (2009). Testing the effectiveness of Content and Language Integrated Learning in foreign language contexts: The assessment of English pronunciation. In Y. Ruiz de Zarobe, & R. M. Jiménez Catalán (Eds.), *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 63- 80). Bristol: Multilingual Matters.

Galton, M. J., Croll, P., & Simon, B. (1980). *Inside the primary classroom.* London: Routledge and Kegan Paul.

García Mayo, M., & Basterrechea, M. (2017). CLIL and SLA: Insights from an interactionist perspective. In A. Llinares, & T. Morton (Eds.), *Applied Linguistic Perspectives on CLIL* (pp. 33-50). Amsterdam: John Benjamins.

García Mayo, M., & Garcia, L. M. L. (2003). *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.

Gardner, S., & Rea-Dickins, P. (2001). Conglomeration or chameleon? Teachers' representations of language in the assessment of learners with English as an additional language. *Language Awareness*, 10, 161-177.

Gardner, S., & Rea-Dickins, P. (2002). Focus on language sampling: a key issue in EAL assessment. *Occasional Paper 15*. National Association for Language Development in the Curriculum: NALDIC Publications Group.

Gass, S. M. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, 9, 198-217.

Gass, S. (1997). *Input, interaction, and the second language learner.* Mahwah, NJ: Lawrence Erlbaum.

Gass, S., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, 19, 3-17.

Gass, S., & Selinker, L. (2001). *Second Language Acquisition: An introductory course.* Mahwah NJ: Lawrence Erlbaum Associates

Gass, S. M., & Varonis, E. (1994). Input, interaction, and second language production. *Studies in Second Language Acquisition*, 16, 283-302.

Gassner, D., & Maillat, D. (2006). Spoken competence in CLIL: A pragmatic take on recent Swiss data. In C. Dalton-Puffer & T. Nikula (Eds.), *Current Research on CLIL. VIEWZ — Vienna English Working Papers*, 15(3), 15-22.

Gearhart, M., Nagashima, S., Pfotenhauer, J., Clark, S., Schwab, C., Vendlinski, T., Osmundson, E., Herman, J., & Bernbaum, D. J. (2006). Developing Expertise With Classroom Assessment in K-12 Science: Learning to Interpret Student Work. Interim Findings From a 2-Year Study. *Educational Assessment*, 11(3-4), 237-263.

Genesee, F. (1987). *Learning through two languages: studies of immersion and bilingual education*. Cambridge, Mass.: London: Newbury House.

Genesee, F. (1994). *Integrating language and content: Lessons from immersion [Educational practice reports no. 11]*. Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning.

Genesee, F. (2004). What do we know about bilingual eduction for majority language students? In T- K- Bhatia, & W. Ritchie (Eds.), *Handbook Of Bilingualism And Multicuturalism* (pp. 547-576). Malden, MA: Blackwell.

Gibbons, P. (1998). Classroom talk and the learning of new registers in a second language. *Language and Education*, 12, 99-118.

Gibbons, P. (2003). Mediating language learning: teacher interactions with ESL students in a content based classroom. *TESOL Quarterly*, 37(2), 247-273.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Falmer Press.

Goffman, E. (1981). *Forms of talk.* Philadelphia, PA: University of Pennsylvania.

Guskey, T. R. (1985). Staff development and teacher change. *Education Leadership*, 42(7), 57-60.

Gutierrez, K. (1994). How talk, context, and script shape context for learning: A cross-case comparison of journal sharing. *Linguistics and Education*, 5, 335-365.

Hall, J. K. (1998). Differential teacher attention to learner utterances: The construction of different opportunities for learning in the IRF. *Linguistics and Education, 9*, 287-311.

Hall, J. K., & Walsh, M. (2002). Teacher-student interaction and language learning. *Annual Review of Applied Linguistics*, 22, 186-203.

Hall, J. K., Webber, B., Varley, S., & Young, V. (1997). A study of teacher assessment at key stage 1. *Cambridge Journal of Education*, 27(1), 107-122.

Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

Hammond, J., & Gibbons, P. (2005). Putting scaffolding to work: The contribution of scaffolding in articulating ESL education. *Prospect*, 20(1), 6-30.

Han, Z.-H. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36, 543-72.

Haneda, M. (2005). Functions of triadic dialogue in the classroom: Examples for L2 research. *The Canadian Modern Language Review*, 62, 313-333.

Haneda, M., & Wells, G. (2008). Learning an additional language through dialogic inquiry. *Language and Education*, 22(2), 114-136.

Hardman, F., Smith, F., & Wall, K. (2003). Interactive whole class teaching' in the National Literacy Strategy. *Cambridge Journal of Education*, 33(2), 197-215.

Harlen, W. (2005). Teachers' summative assessment practices and assessment for learning: Tensions and synergies. *Curriculum Journal*, 16(2), 207-223.

Harlen, W., & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning (EPPI-Centre Review). *Research Evidence in Education Library. Issue 1*. London: EPPI-Centre, Social Science Research Unit, Institute of Education. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=61 (accessed March 20, 2010).

Harlen, W., & Deakin-Crick, R. (2003). Testing and motivation for learning. *Assessment in Education*, 10(2), 169-208.

Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education*, 4(3), 365-379.

Harlen, W., & Winter, J. (2004). The development of assessment for learning: learning from the case of science and mathematics. *Language Testing*, 21(3), 390-408.

Harley, B., Allen, P., Cummins, J. and Swain, M. (Eds.) (1990). *The Development of Second Language Proficiency*. Cambridge: Cambridge University Press.

Harley, B. (1993). Instructional strategies and SLA in early French immersion. *Studies in Second Language Acquisition*, 15, 245-260.

Harrison, C., & Howard, S. (2009). *Inside the Primary black box*. London, UK: GL Assessment.

Harrison & Swaffield (2003). Assessment for learning: the KMOFAP Project. In *Whither Assessment*. London: Qualifications and Curriculum Authority.

Hatch, E. M. (Ed.). (1978). *Second language acquisition*. Rowley, MA: Newbury House.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.

Havranek, G., & Cesnik, H. (2001). Factors affecting the success of corrective feedback. In S. Foster-Cohen, & A. Nizegorodzew (Eds.), *EUROSLA Yearbook* Vol. 1. (pp. 99-122).

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140-145.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment.* Washington, DC: Chief Council of State School Officers.

Heritage, M. (2010). *Formative Assessment and Next-Generation Assessment Systems: Are We Losing an Opportunity?* Paper prepared for the Council of Chief State School Officers.

Hickey, D. T., Zuiker, S., Taasoobshirazi, G., Schafer, M., & Michael, M. (2006). Balancing varied assessment functions to attain systemic validity: Three is the magic number. *Studies in Educational Evaluation*, 32, 180-201.

Hiebert, J. (1999). Relationships between research and the NCTM standards. *Journal for research in mathematics education*, 30(1), 3-19.

Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: a comparison. *The Curriculum Journal*, 16(2), 153-176.

Hönig, I. (2010). *Assessment in CLIL: Theoretical and Empirical Research.* Saarbrucken: VDM Verlag.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels.* Champaign, IL: National Council of Teachers of English.

Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form of *-te i (ru)* by learners of Japanese as a foreign language. *Language Learning*, 54, 311-394.

Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects of L2 development. *Studies in Second Language Acquisition*, 25, 1-36.

Jacoby, S., & Ochs, E. (1995). Co-construction: An introduction. *Research on Language and Social Interaction*, 28(3), 171-183.

Jäppinen, A.-K. (2005). Thinking and Content Learning of Mathematics and Science as Cognitional Development in Content and Language Integrated Learning (CLIL): Teaching

Through a Foreign Language in Finland. *Language & Education: An International Journal*, 19, 148-169.

Jarvis, J., & Robinson, M. (1997). Analysing educational discourse: an exploratory study of teacher response and support to pupils learning. *Applied Linguistics*, 18(2), 212-228.

Jexenflicker, S., & Dalton-Puffer, C. (2010). The CLIL differential: Comparing the writing of CLIL and non-CLIL students in higher colleges of technology. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 169-190). Amsterdam: John Benjamins.

Jiménez Catalán, R., Ruiz de Zarobe, Y., & Cenoz, J. (2006). Vocabulary Profiles of English Foreign Language Learners in English as a Subject and as a Vehicular Language. *VIEWS (Vienna English Working Papers)*, 15(3), 23-27.

Kerry, T. (1989). Analysing the cognitive demands made by classroom tasks in mixed-ability classes. In E. C. Wragg (Ed.), *Classroom Teaching Skills* (pp. 163-179). London: Routledge.

Klecan-Aker, J. S., & López, B. (1985). A comparison of T-units and cohesive ties used by first and third grade children. *Language and Speech*, 28(3), 307-315.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Towards the understanding of a double-edge sword. *Current Directions in Psychological Science*, 7, 67-72.

Krashen, S. D. (1980). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. D.(1982) *Principles and practice in second language acquisition*. Oxford: Pergamon.

Krashen, S. D. (1983). Newmark's Ignorance Hypothesis and current second language acquisition theory. In S. M. Gass, & L. Selinker (Eds.), *Language transfer in language learning* (pp. 135-156). Rowley, MA: Newbury House.

Krashen, S. D. (1985). *The Input Hypothesis: Issues and implications.* London: Longman.

Krashen, S. D. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45-77). London: Academic Press.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218.

Kupetz, M. (2011). Multimodal resources in students' explanations in CLIL interaction. *Novitas- Royal*, 5(1), 121-142.

Lasagabaster, D. (2008). Foreign language competence in content and language integrated learning. *Open Applied Linguistics Journal*, 1, 31-42.

Lasagabaster, D., & Ruiz de Zarobe, Y. (2010). *CLIL in Spain: implementation, results and teacher training*. UK: Cambridge Scholars Publishing.

Lasagabaster, D., & Sierra, J. M. (2009). Language attitudes in CLIL and traditional EFL classes. *International CLIL Research Journal*, 1(2), 4-17.

Lantolf, J. P., & Thorne, S. L. (2006). Sociocultural theory and second language acquisition. In B. van Patten, and J. Williams (Eds.), *Theories in second language acquisition* (pp. 201-224). Mahwah, NJ: Erlbaum.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute-by-minute and day-by-day. *Educational Leadership*, 63(3), 18-24.

Lee, Y. (2007). Third turn position in teacher talk: contingency and the work of teaching. *Journal of Pragmatics*, 39, 180-206.

Leeman, J. (2000). *Towards a new classification of input: An empirical study of the effect of recasts, negative evidence and enhanced salience on L2 development.* Unpublished Ph.D. dissertation. Georgetown university, Washington D.C.

Leeman, J. (2003). Recasts and L2 development: Beyond negative evidence. *Studies in Second Language Acquisition*, 25, 37-63.

Lemke, J. L. (1982). *Classroom communication of science* (Final Report to the National Science Foundation). New York: Brooklyn College, City University of New York (ERIC Document Reproduction Service No. ED 222 346).

Lemke, J. L. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex.

Leung, C. (1999). Teachers' response to linguistic diversity. In A. Tosi, & C. Leung (Eds.), *Rethinking language education: From a monolingual to a multilingual perspective* (pp. 225-240). London: Centre for Information on Language Teaching and Research, the National Centre for Languages.

Leung, C. (2001). English as an additional language: Distinct language focus or diffused curriculum concerns? *Language and Education*, 15(1), 33-55.

Leung, C. (2004). Developing formative assessment: knowledge, practice and change. *Language Assessment Quarterly*, 1(1), 19-41.

Leung, C. (2005a). Language and content in bilingual education. *Linguistics and Education,* 16, *238-252.*

Leung, C. (2007). Dynamic assessment: assessment *for* and *as* teaching? *Language Assessment Quarterly*, 4(3), 257-278.

Leung, C. (2010). English as an additional language: Learning and participating in mainstream classrooms. In P. Seedhouse, S. Walsh, & C. Jenks (Eds.), *Conceptualising learning in applied linguistics* (pp. 182-205). Basingstoke: Palgrave Macmillan.

Leung, C., & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: language testing and assessment. *TESOL Quarterly*, 40(1), 211-234.

Leung, C., & Mohan, B. A. (2004). Teacher formative assessment and talk in classroom contexts: assessment as discourse and assessment of discourse. *Language Testing*, 21(3), 335-359.

Leung, C., & Mohan, B. A. (in press). Teacher formative assessment and talk in classroom contexts-as- sessment as discourse and assessment of discourse. *Language Testing*.

Leung, C**.,** & Rea-Dickins, P. (2007). Teacher assessment as policy instrument - contradictions and capacities. *Language Assessment Quarterly*, 4(1), 6-36.

Leung, C., & Teasdale, A. (1997). Raters' understanding of rating scales as abstracted concept and as instruments for decision-making. *Melbourne Papers in Language Testing*, 6, 45-70.

Lightbown, P. (1998). The importance of timing in focus on form. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 177-196). Cambridge: Cambridge University Press.

Lightbown, P. M. (2000). Anniversary article: Classroom SLA research and second language teaching. *Applied Linguistics*, 21, 431-62.

Lightbown, P. M., & Spada, N. (1990). Focus-on-form and corrective feedback incommunicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition*, 12, 429-448.

Lightbown, P. M., & Spada, N. (2006). *How Languages are Learned*. Oxford: Oxford University Press.

Lim Falk, M. (2008). *Svenska i engelskspråkig skolmiljö. Ämnesrelaterat språkbruk i två gymnasieklasser*. (Swedish in an English-language School Environment. Subject-based Language Use in Two Upper Secondary Classes.)

Llinares, A. (2015). Integration in CLIL: A proposal to inform research and successful pedagogy. *Language, Culture and Curriculum*, 28(1), 58-73.

Llinares, A., & Dafouz, E. (2010). Content and language integrated programmes in the Madrid region: Overview and research findings. In D. Lasagabaster, & Y. Ruiz de Zarobe, (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training* (pp. 95-114). UK: Cambridge Scholars Publishing.

Llinares, A., & Dalton-Puffer, C. (2015). The role of different tasks in CLIL students' use of evaluative language. *System*, 54, 69-79.

Llinares, A., & Lyster, R. (2014). The influence of context on patterns of corrective feedback and learner uptake: A comparison of CLIL and immersion classrooms. *The Language Learning Journal*, 42(2), 181-194.

Llinares, A., & Morton, T. (2010). Historical explanations as situated practice in content and language integrated learning. *Classroom Discourse*, 1(1), 46-65.

Llinares, A., & Morton, T. (2012). Social perspectives on interaction and language learning in CLIL classrooms. In E. Alcón Soler, & P. Safont Jordá (Eds.), *Discourse and learning across L2 instructional contexts*. Amsterdam: Rodopi.

Llinares, A., & Morton, T. (2017). *Applied linguistics perspectives on CLIL*. Amsterdam: John Benjamins.

Llinares, A., Morton, T., & Whittaker, R. (2012). *The roles of language in CLIL*. Cambridge: CUP.

Llinares, A., & Nikula, T. (2016). *Conceptualising Integration in CLIL and Multilingual Education*. Clevedon: Multilingual Matters.

Llinares, A., & Pascual, I. (2015). A genre approach to the effect of academic questions on CLIL students' language production. *Language and Education*, 29(1), 15-30.

Llinares, A., & Pastrana, A. (2013). CLIL students' pragmatic development across classroom activities and educational levels. *Journal of Pragmatics*, 59(A), 81-92.

Llinares, A., & Whittaker, R. (2007). Talking and writing in a foreign language in CLIL contexts: a linguistic analysis of secondary school learners of geography and history. *Revista Española de Lingüística Aplicada, Volumen Monográfico*, 83-94.

Llinares, A., & Whittaker, R. (2009). Teaching and learning history in secondary CLIL classrooms: from speaking to writing. In E. Dafouz, & M. C. Guerrini (Eds.), *CLIL across Educational Levels* (pp. 73-89). Madrid: Santillana Educación S.L./Richmond Publishing.

Llinares, A., & Whittaker, R. (2010). Writing and speaking in the history class: A comparative analysis of CLIL and first language contexts. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 125-143). Amsterdam: John Benjamins.

Lochtmann, K. (2007). Die mündliche Fehlerkorrektur in CLIL und im traditionellen Fremdsprachenunterricht: Ein Vergleich. [Oral corrective feedback in CLIL and traditonal foreign language teaching: A comparison]. In C. Dalton-Puffer, & U. Smit (Eds.), *Empirical perspectives on CLIL classroom discourse* (pp. 119-138). Frankfurt, Germany: Peter Lang.

Long, M. H. (1977). Teacher feedback on learner error: Mapping cognitions. In H. D. Brown, C. A. Yorio, & R. H. Crymes (Eds.), *On TESOL '77* (pp. 278-293).Washington: TESOL.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4, 126-141.

Long, M. H. (1985). Input and second language acquisition theory. In S. M. Gass, & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 377-393). Rowley, MA: Newbury House.

Long, M. H. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly*, 24, 649-666.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamins.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie, & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Long, M. (2007). Recasts in SLA: The story so far. In M. H. Long (Ed.), *Problems in SLA* (pp. 75-116). Mahwah, NJ: Laurence Erlbaum.

Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal*, 82, 357-71.

Long, M. H., & Robinson, H. (1998). Focus on form: Theory, research, process. In C. Doughty, & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 15-41). Cambridge: Cambridge University Press.

Long, M. H., & Sato C. J. (1983). Classroom foreigner talk discourse: forms and functions of teachers' questions. In H. Seliger, & M.H. Long (Eds.), *Classroom Oriented Research in Second Language Acquisition* (pp. 268-285). Cambridge: Newbury House Publishers.

Lorenzo, F., Casal, S., & Moore, P. (2005). *Orientaciones para la elaboración del currículo integrado de las lenguas en los centros bilingües*. Seville, Spain: Consejería de Educación (Junta de Andalucía).

Lorenzo, F., Casal, S., & Moore, P. (2009). The effects of content and language integrated learning in European education: Key findings from the Andalusian Bilingual Sections evaluation project. *Applied Linguistics*, 31(3), 418-442.

Lundgren, U. P. (1977). *Model analysis of pedagogical processes*. Lund: Liber/Gleerup.

Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18, 351-372.

Lyster, R. (1987). Speaking immersion. *The Canadian Modern Language Review*, 43, 701-717.

Lyster, R. (1998a). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20, 51-81.

Lyster, R. (1998b). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, 48, 183-218.

Lyster, R. (1998c). Immersion pedagogy and implications for language teaching. In J. Cenoz, & F. Genesee (Eds.), *Beyond bilingualism: Multilingualism and multilingual education* (pp. 64-95). Clevedon, UK: Multilingual Matters.

Lyster, R. (1998d). Form in immersion classroom discourse: In or out of focus? *Canadian Journal of Applied Linguistics*, 1, 53-82.

Lyster, R. (2002). The importance of differentiating negotiation of form and meaning in classroom interactions. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An Integrated View of Language Development. Papers in Honor of Henning Wode* (pp. 381-398). Trier: WVT Wissenschaftlicher Verlag.

Lyster, R. (2002a). Negotiation in immersion teacher-student interaction. *International Journal of Educational Research*, 37(3-4), 237-53.

Lyster, R. (2004a). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 19, 37-66.

Lyster, R. (2004b). Research on form-focused instruction in immersion classrooms: implications for theory and practice. *Journal of French Language Studies*, 14(3), 321- 341.

Lyster, R. (2007). *Learning and teaching languages through content. A counterbalanced approach*. Amsterdam: John Benjamins.

Lyster, R., & Ballinger, S. (2011). Content-based language teaching: Convergent concerns across divergent contexts. *Language Teaching Research*, 15(3), 279-288.

Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, 28(2), 269-300.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. Negotiation of form in communicative classrooms. *Studies in Second language Acquisition*, 19, 37-66.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46(1), 1-40.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21, 557-87.

Mackey, A. (2000). Interactional feedback in the L2 classroom: learner's perceptions and developmental outcomes. Plenary Address, Penn TESOL East Conference.

Mackey, A. (2006). Feedback, noticing, and instructed second language learning. *Applied Linguistics*, 27, 405-430.

Mackey, A. (in press). The relationship between conversational interaction and second language acquisition. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition: A Collection of Empirical Studies* (Chapter One). Oxford: Oxford University Press.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22, 471-497.

Mackey, A., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System*, 30, 459-77.

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads. *Language Learning*, 53, 35-66.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal*, 82, 338-56.

Mackey, A., & Silver, R. E. (2005). Interactional tasks and English L2 learning by immigrant children in Singapore. *System*, 33, 239-60.

Maillat, D. (2010). The pragmatics of L2 in CLIL. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 39-60). Amsterdam: John Benjamins.

Maillat, D. & Serra, C. (2009). Immersive Education and Cognitive Strategies: Can the Obstacle Be the Advantage in a Multilingual Society? In L. Gajo, & D. Moore (Eds.), *French Voices on Plurilingualism and Pluriculturalism : Theory, Significance and Perspectives. Special issue of International Journal of Multilingualism.*

Markee, N., & Kasper, G. (2004). Classroom talks: An introduction. *The Modern Language Journal*, 88(4), 491-500.

Martín Úriz, A., & Whittaker, R. (Eds.) (2005). *La composición como comunicación: una experiencia en las aulas de lengua inglesa en Bachillerato*. Madrid: Universidad Autónoma Ediciones.

Marsh, D. (Ed.) (2002). *CLIL/EMILE - The European Dimension: Actions, Trends and Foresight*. Potential Public Services Contract DG EAC: European Commission.

Marzano, R. J. (2010). *Formative Assessment & Standards-Based Grading.* Bloomington: Marzano Research Laboratory.

McCabe, A. & Whittaker, R. (2016). Genre and appraisal in CLIL history texts: Developing the voice of the historian. In A. Llinares, & T. Morton (Eds.), *Applied Linguistics Perspectives on CLIL* (pp. 105-124). Amsterdam: John Benjamins.

McCallum, B., Gipps, C., McAlister, S., & Brown, M. (1995). National Curriculum assessment: emerging models of teacher assessment in the classroom. In H. Torrance (Ed.), *Evaluating authentic assessment* (pp. 57-87). Buckingham: Open University Press.

McCormick, D., & Donato, R. (2000). The discourse of teacher questions as scaffolding in an integrated ESL classroom. In J. K. Hall, & L. S. Verplaeste (Eds.), *Second and foreign language learning through classroom interaction* (pp. 183-201). Mahwah, NJ: Erlbaum.

McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, 27, 79-103.

McKay, P. (2006). *Assessing young language learners.* Cambridge: Cambridge University Press.

McNamara, T. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics*, 18(4), 446-66.

McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing* 18(4), 333-349.

Mehan, H. (Ed.). (1979). *Learning Lessons: Social Organization in the Classroom.* Harvard: Harvard University Press.

Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL*. Oxford, UK: Macmillan.

Menegale, M. (2011). Teacher questioning in CLIL lessons: How to enhance teacher-students interaction. In C. Escobar Urmeneta, N. Evnitskaya, E. Moore, & A. Patiño (Eds.), *AICLE-CLIL-EMILE: Educació plurlingüe. Experiencias, research & polítiques* [Plurilingual education. Experiences, research & politics] (pp. 83-99). Bellaterra, Spain: Universitat Autonoma de Barcelona.

Mercer, N. (1992). Talk for teaching and learning. In K. Norman (ed.), *Thinking voices: The work of the National Oracy Project* (PP. 215-223). London: Hodder & Stoughton.

Mercer, N. (1995). *The guided construction of knowledge.* Clevedon: Multilingual Matters.

Mercer, N. (2000). *Words and minds: how we use language to think together*. London: Routledge.

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004) Reasoning as a scientist: ways of helping children to use language to learn science. *British Educational Research Journal*, 30, 359 - 77.

Met, M. (1994) Teaching content through a second language. In F. Genessee (Ed.), *Educating Second Language Children* (pp. 159-182). Cambridge: CUP.

Met, M. (1998). Curriculum decision-making in content-based language teaching. In J. Cenoz, & F. Genesee (Eds.), *Beyond bilingualism: Multilingualism and multilingual education* (pp. 35-63). Clevedon: Multilingual Matters.

Mewald, C. (2007). A comparison of oral foreign language performance of learners in CLIL and mainstream classes at lower secondary level in Lower Austria. In C. Dalton-Puffer, & U. Smit (Eds.), *Empirical perspectives on CLIL classroom discourse* (pp. 139-173). Frankfurt: Peter Lang.

Meyer, W. (1982). Indirect communication about perceived ability estimates. *Journal of Educational Psychology*, 74, 888-897.

Meyer, W., Bachmann, U., Hempelmann, M., Ploger, F., & Spiller, H. (1979). The informational value of evaluation behavior: Influences of praise and blame in perceptions of ability. *Journal of Educational Psychology*, 71, 259-268.

Mohan, B., & Beckett, G.H. (2003). A functional approach to research on content-based language learning: Recasts in causal explanations. *The Modern Language Journal*, 87(3), 421- 32.

Mohan, B. A., Leung, C., & Slater, T. (2010). Assessing language and content: a functional perspective. In A. Paran, & L. Sercu, L. (Eds.) *Testing the untestable in language education* (pp. 217-240). Bristol: Multilingual matters.

Mohan, B. A., & Slater, T. (2004). The evaluation of causal discourse and language as a resource for meaning. In J. A. Foley (Ed.), *Language, education, and discourse: Functional approaches* (pp. 255-269). New York: Continuum.

Moore, D. (2002). Case study: Code-switching and learning in the classroom. *International Journal of Bilingual Education and Bilingualism*, 5, 279-293.

Moore, P. (2011). Collaborative interaction in turn-taking: a comparative study of European bilingual (CLIL) and mainstream (MS) foreign language learners in early secondary education. *International Journal of Bilingual Education and Bilingualism*, 14(5), 531-549.

Moore, P., & Nikula, T. (2016). Translanguaging in CLIL Clasrooms. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising Integration in CLIL and Multilingual Education* (pp. 211-234). Bristol: Multilingual Matters.

Mori, J. (2002). Task design, plan, and development of talk-in-interaction: An analysis of a small group activity in a Japanese language classroom. *Applied Linguistics*, 23, 323-347.

Morris, F. (2002). Negotiation and recasts in relation to error types and learner repair. *Foreign Language Annals*, 35, 395-404.

Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Maidenhead: Open University Press.

Morton, T. (2012). Classroom talk, conceptual change and teacher reflection in bilingual science teaching. *Teaching and Teacher Education*, 28(1), 101-110.

Morton, T & Llinares, A. (2016). Students' use of evaluative language in L2 English to talk and write about history in a bilingual education programme. *International Journal of Bilingual Education and Bilingualism*, 1-13.

Musumeci, D. (1996). Teacher-learner negotiation in content-based instruction: Communication at Cross-Purposes? *Applied Linguistics*, 17(3), 286-325.

Nassaji, H., & Wells, G. (2000). What's the use of triadic dialogue? An investigation of teacher-student interaction. *Applied Linguistics*, 21(3), 376-406.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, 155-175.

Navés, T. (2002). Successful CLIL Programmes. In T. Navés, C. Munõz, & M. Pavesi (Eds.), *The CLIL Professional Development Course.* Milan: Ministero della´Istruzione della´Università e della Ricerca.

Navés, T. (2009). Effective Content and Language Integrated Learning (CLIL) programmes. In Y. Ruiz de Zarobe, & R. M. Jiménez Catalán (Eds.), *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 22-40). Bristol: Multilingual Matters.

Netten, J. (1991). Towards a more language oriented second language classroom. In L. Malavé, & G. Duquette (Eds.), *Language, culture and cognition* (pp. 284-304). Clevedon, UK: Multilingual Matters.

Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: Working for cognitive change in school.* Cambridge: Cambridge University Press.

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, 51, 719-58

Nikula, T. (2002). Teacher talk reflecting pragmatic awareness: A look at EFL and content-based classroom settings. *Pragmatics*, 12(4), 447-467.

Nikula, T. (2005). English as an object and tool of study in classrooms: Interactional effects and pragmatic implications. *Journal Linguistics and Education*, 16, 1, 27-58.

Nikula, T. (2007a). The IRF pattern and space for interaction: observations on EFL and CLIL classrooms. In C. Dalton-Puffer, & U. Smit (Eds.), *Empirical perspectives on CLIL classroom discourse* (pp. 179-204). Frankfurt: Peter Lang.

Nikula, T. (2007b). Speaking English in Finnish content-based classrooms. *World Englishes*, 26(2), 206-223.

Nikula, T. (2008). Learning pragmatics in content-based classrooms. In E. A. Soler, & A. Martinez-Flor (Eds.), *Investigating Pragmatics in Foreign Language Learning, Teaching, and Testing* (pp. 94-113). Clevedon: Multilingual Matters.

Nikula, T. (2012). On the role of peer discussions in the learning of subject-specific language use in CLIL. In E. A. Soler, & M.-P. Safont-Jordà (Eds.), *Discourse and language learning across L2 instructional settings* (pp. 133-153). Utrecht Studies in Language and Communication, 24. Amsterdam: Rodopi.

Nikula, T., Dafouz, E., Moore, P., & Smit, U. (Eds.). (2016). *Conceptualising Integration in CLIL and Multilingual Education*. Multilingual Matters.

Nikula, T., Dalton-Puffer, C., & Llinares, A. (2013). CLIL classroom discourse: Research from Europe. *Journal of Immersion and Content-Based Language Education*, 1(1), 70-100.

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2009). Syntactic Development in Adolescents with a History of Language Impairments: A Follow-Up Investigation. *American Journal of Speech-Language Pathology*, 18, 241-251.

Nitko, A. J. (1989). Designing tests that are integrated with assessment. In R. L. Linn (Ed.), *Educational measurement.* 3rd edn. New York: American Council on Education / Macmillan.

Nitko, A.J. (1994). *A model for developing curriculum-driven criterion-referenced and norm-referenced national examinations for certification and selection of students.* A paper presented the 2nd International Conference on Educational Evaluation and Assessment of the Association for the Study of Educational Evaluation in Southern Africa, Pretoria, South Africa, July.

Nobuyoshi, J., & Ellis, R. (1993). Focused communication tasks and second language acquisition. *English Language Teaching Journal*, 47, 203-210.

Nystrand, M. (1997). Dialogic instruction: When recitation becomes conversation. In M. Nystrand, A. Gamoran, P. Kachur, & C. Prendergast (Eds.), *Opening dialogue: understanding the dynamics of language and learning in the English classroom* (pp. 1-29). New York: Teachers College Press.

Nystrand, M., & Gamoran, A. (1991). Student engagement: When recitation becomes conversation. In H. C. Waxman, & H. J. Walberg (Eds.), *Effective teaching: Current research* (pp. 257-276). Berkeley, CA: McCutchan.

Nystrand, M., & Gamoran, A. (1997). The big picture: Language and learning in hundreds of English lessons. In M. Nystrand, A. Gamoran, P. Kachur, & C. Prendergast (Eds.), *Opening dialogue: understanding the dynamics of language and learning in the English classroom* (pp. 30-74). New York: Teachers College Press.

Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue: understanding the dynamics of language and learning in the English classroom.* New York: Teachers College Press.

O'Connor, M. C., & Michaels, S. (1993). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly*, 24, 318-335.

O'Donnell, M. (2011). *UAM CorpusTool* (http://www.wagsoft.com/CorpusTool).

Ohta, A. (2001). *Second language acquisition processes in the classroom: Learning Japanese.* Lawrence Erlbaum Associates.

Olina, Z., & Sullivan, H. J. (2004). Student self-evaluation, teacher evaluation, and learner performance. *Educational Technology Research and Development*, 52(3), 5-22.

Oliver, R. (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition*, 17, 459-481.

Oliver, R. (2000). Age differences in negotiation and feedback in classroom and pairwork. *Language Learning*, 50, 119-51.

Oliver, R. and Mackey, A. (2003). Interactional context and feedback in child ESL classrooms. *Modern Language Journal*, 87(4), 519-543.

O'Malley, J. M., & Valdez Pierce, L. (1996). *Authentic assessment for English language learners.* Reading, MA: Addison-Wesley.

Ortega, L. & Long, M. H. (1997). The effects of models and recasts on the acquisition of object topicalization and adverb placement by adult learners of Spanish. *Spanish Applied Linguistics*, 1, 65-86.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36, 573-595.

Pascual, I. (2010). Teachers' questions in CLIL contexts. *VIEWZ (Vienna English Working Papers)*, 19(3), 65-71.

Pascual, I. & Basse, R. (2017). The use of meta-questions in primary bilingual CLIL classrooms: Building an Assessment for Learning discourse. In A. Llinares, & T. Morton (Eds.), *Applied Linguistics Perspectives on CLIL* (pp. 221-235). Amsterdam: John Benjamins.

Pastrana, A. (2010). Language functions in CLIL classrooms: Students' oral production in different classroom activities. *VIEWZ*, 19(3), 72-82.

Pekarek Doehler, S. (2010). Conceptual changes and methodological challenges: on language, learning and documenting learning in conversation analytic SLA research. In P. Seedhouse, S. Walsh, & C. Jenks (Eds.), *Conceptualising learning in applied linguistics* (pp. 105-126). Basingstoke: Palgrave MacMillan.

Pekarek Doehler, S., & Ziegler, G. (2007). Doing language, doing science and the sequential organization of the immersion classroom. In Z. Hua, P. Seedhouse, L. Wei, & V. Cook (Eds.), *Language learning and teaching as social inter-action* (pp. 72-86). Basingstoke, UK: Palgrave MacMillan.

Pérez-Vidal, C. (2009). The integration of content and language in the classroom: A European approach to education (the second time around). In E. Dafouz, & M. C. Guerrini (Eds.), *CLIL across Educational Levels* (pp. 3-16). Madrid: Santillana Educación S.L./Richmond Publishing.

Perrenoud, P. (1991) Towards a pragmatic approach to formative evaluation. In P. Weston (Ed.), *Assessment of pupil achievement* (pp. 79-101). Amsterdam, Swets & Zeitlinger.

Philip, J. (2003). Constraints on "noticing the gap": Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, 25, 99- 126.

Pica, T. (1994). Research on negotiation: what does it reveal about second- language learning conditions, processes and outcomes? *Language Learning*, 44, 493-527.

Pica, T. (1995). Teaching language and teaching language learners: The expanding roles and expectations of language teachers in communicative, content-based classrooms. In J. E. Alatis, C. A. Straehle, B. Galenberger, & M. Ronkin (Eds.), *Linguistics and the education of language teachers: Ethnolinguistic, psycholinguistic and sociolinguistic aspects* (pp. 378-397). Washington DC: Georgetown University Round Table on Languages and Linguistics.

Pica, T. (2002). Subject-matter content: how does it assist the interactional and linguistic needs of classroom language learners? *The Modern Language Journal*, 86(1), 1-19.

Pica, T., Holliday, L., Lewis, N., & Mortgenhaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, 11, 63-90.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communicative tasks for second language instruction and research. In S. Gass, & G. Crookes (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 9-34). Clevedon, UK: Multilingual Matters.

Pica, T., Young, R., & Doughty, C. (1987). The impact of inter- action on comprehension. *TESOL Quarterly*, 21, 737-58.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 10, 52-79.

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching and assessing. *Theory into Practice*, 41(4), 219-225.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. New York: Springer.

Polias, J. (2003). *ESL scope and scales.* Adelaide, South Australia: Department of Education and Children's Services.

Pryor, J., & Torrance, H. (1996). Teacher-pupil interaction in formative assessment: assessing the work or protecting the child? *The Curriculum Journal*, 7, 205-226.

Ramaprasad , A. (1983). On the definition of feedback. *Behavioural Science*, 28(1), 4-13.

Rawlins, P. (2010). Student participation in formative assessment for NCEA. *New Zealand Journal of Educational Studies*, 45(1), 3-16.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21, 249-259.

Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics*, 16(2), 163-188.

Rea-Dickins, P. (2008). Classroom-based language assessment. In E. Shohamy, & N. H. Hornberger (Eds.), *Encyclopedia of language and education. Vol. 7: Language testing and assessment* (2nd ed., pp. 257-271). New York: Springer Science + Business.

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: disentangling the construct of formative assessment. *Language Testing*, 17(2), 215-243.

Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research*, 51, 237-245.

Rex, L., & McEachen, D. (1999). "If anything is odd, inappropriate, confusing, or boring, it's probably important": The emergence of inclusive academic literacy through English classroom discussion practices. *Research in the Teaching of English*, 34(1), 65-129.

Riesco Bernier, S. (2007). *Modelling the discourse-grammar interface of EFL Pre-school teacher talk (native and non-native): proposal of a system network enabling the systematic analysis of regulatory functions*. Doctoral Thesis. Universidad Autónoma de Madrid.

Roberts, M. A. (1995). Awareness and the efficacy of error correction. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 163-82). University of Hawai'i Press.

Rogoff, B. (1994). Developing understanding of the idea of Communities of Learners. *Mind, Culture and Activity*, 1(4), 209-229.

Romero Trillo, J., & Llinares, A. (2001). Communicative Constraints in EFL Pre-school Settings: a Corpus-Driven Approach. *International Journal of Corpus Linguistics*, 6(1), 27-46.

Rosenshine, B. (1976). Classroom instruction. In N. L. Gage (Ed.), *The psychology of teaching methods. The seventy-fifth yearbook of the National Society for the Study of Education* (pp. 335-271). Chicago: University of Chicago.

Rowe, M. B. (1974). Wait time and rewards as instructional variables, their influence on language, logic and fate control. *Journal of Research in Science Teaching*, 11, 81-94.

Ruiz de Zarobe, Y. (2008). CLIL and foreign language learning: A longitudinal study in the Basque Country. *International CLIL Research Journal*, 1, 60-73.

Ruiz de Zarobe, Y. (2010). Written production and CLIL: An empirical study. In C. Dalton Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 191-210). Amsterdam: John Benjamins.

Ruiz de Zarobe, Y., & Jiménez Catalán, R. M. (Eds.) (2009). *Content and language integrated learning: evidence from research in Europe*. Bristol: Multilingual Matters.

Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3/4), 205 - 235.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57-84.

Ruiz- Primo, M. A., Shalvelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, 39(5), 369-393.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.

Sadler, R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education*, 5(1), 77-84.

Salomone, A. (1992). Student-teacher interactions in selected French immersion classrooms. In E. Bernhardt (Ed.), *Life in language immersion classrooms* (pp. 97-109). Clevedon, UK: Multilingual Matters.

Sato, C. (1986). Conversation and interlanguage development: Rethinking the connection. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 23-48). Rowley, MA: Newbury House

Savignon, S. (2004). Communicative Language Teaching. In M. Byram (Ed.), *The Routledge encyclopaedia of language teaching and learning* (pp. 124-129). London: Routledge.

Schachter, J. (1983). Nutritional needs of language learners. In M. Clarke, & J. Handscombe (Eds.), *On TESOL '82: Pacific perspectives on language learning and teaching* (pp. 175-89). Washington, DC: TESOL.

Schindelegger, V. (2010). *The IRF sequence in CLIL and EFL classrooms* (Unpublished MA Thesis). University of Vienna, Austria. Retrieved on 1 November 2012, from http://othes.univie.ac.at/4960/

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics approach*. Mahwah, NJ: Lawrence Erlbaum.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-58.

Schuitemaker-King, J. (2012). *Teachers' strategies in providing opportunities for second language development* (Unpublished PhD dissertation). University of Groningen, The Netherlands.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation, Volume I* (pp. 39-83). Chicago, IL: Rand McNally.

Sebatane, E. M. (1998). Assessment and classroom learning: a response to Black and Wiliam. *Assessment in Education: Principles, Policy and Practice*, 5(1), 123-130

Seedhouse, P. (1997). The case of the missing ‘no’: The relationship between pedagogy and interaction. *Language Learning*, 47(3), 547-583.

Seedhouse, P., & Walsh, S. (2010). Learning a second language through classroom interaction. In P. Seedhouse, S. Walsh, & C. Jenks (Eds.),*Conceptualising learning in applied linguistics* (pp. 127-146). Basingstoke: Palgrave Macmillan.

Seikkula-Leino, J. (2007). CLIL learning: achievement levels and affective factors. *Language and Education*, 21(4), 328-341.

Serra, C. (2007). Assessing CLIL at primary school: a longitudinal study. *The International Journal of Bilingual Education and Bilingualism*, 10(5), 582-602.

Sharpe, T. (2006). 'Unpacking' scaffolding: Identifying discourse and multimodal strategies that support learning. *Language and education*, 20(3), 211-231.

Sharwood Smith, M. (1987). Comprehension vs. acquisition: Two ways of processing input. *Applied Linguistics*, 8, 237-255.

Shavelson, R. J., Black, P., Wiliam, D., & Coffey, J. (2003). On aligning summative and formative functions in the design of large-scale assessment systems. *Educational Evaluation and Policy Analyses*.

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8, 263-300.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research*, 10(4), 361-392.

Sheen, Y., & Ellis, R. (2011). Corrective feedback in language teaching. In E. Hinkel (Ed.), *Handbook of research in second language learning and teaching*. *Volume II* (pp. 593-610). London and New York: Routledge.

Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1066-1101).Washington, DC: American Educational Research Association.

Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, 63(3), 66-71.

Shinn, M. R., & Hubbard, D. D. (1992). Curriculum-based measurement and problem-solving assessment: Basic procedures and outcomes. *Focus on exceptional children*, 24(5), 1-20.

Shohamy, E. (1992). New modes of assessment: the connection between testing and learning. In E. Shohamy, & A. Ronald Walton (Eds.), *Language assessment for feedback: testing and other strategies* (pp. 7-28). USA: Kendall/Hunt Publishing Company.

Silver, R. E. (2000). Input, output, and negotiation: Conditions for second language development. In B. Swierzbin, F. Morris, M. E. Anderson, C. A. Klee, & E. Tarone (Eds.), *Social and Cognitive Factors in Second Language Acquisition: Selected proceedings of the 1999 Second Language Research Forum* (pp. 345-71). Somerville, MA: Cascadilla.

Sinclair, J. M., & Brazil, D. (1982). *Teacher talk*. Oxford: Oxford University Press.

Sinclair, J. M., & Coulthard, M. (1975). *Towards an Analysis of Discourse: The English used by teachers and pupils.* London: Oxford University Press.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Slater, T., & Mohan, B. (2010). Towards systematic and sustained formative assessment of causal explanations in oral interactions. In A. Paran, & L. Sercu (Eds.), *Testing the untestable in language education* (pp. 256-269). Bristol, UK: Multilingual Matters.

Slavin, R. E. (1991). Synthesis of research on cooperative learning. *Educational Leadership*, 48(5), 71-82.

Smit, U. (2010). Conceptualising English as a lingua franca (ELF) as a tertiary classroom language. *Stellenbosch Papers in Linguistics*, 39, 59-74.

Smit, U., & Dafouz, E. (2012). Integrating content and language in higher education. An introduction to English-medium policies, conceptual issues and research practices across Europe. *AILA Review*, 25, 1-12.

Snell, J., & Lefstein, A. (2011). Computer-assisted systematic observation of classroom discourse and interaction. *Working papers on urban language & literacies* 77.

Snow, M. A., Met, M., & Genesee, F. (1989). A conceptual Framework for the Integration of Language and Content in Second/Foreign Language Instruction. *TESOL Quarterly*, 23(2), 201-217.

Snow, M. A., Met, M., & Genesee, F. (1992). A conceptual framework for the integration of language and content instruction. In P. A. Richard-Amato, & M. A. Snow (Eds.), *The multicultural classroom: Readings for content-area teachers* (pp. 27-38). New York: Longman.

Spolsky, B. (1992). The gentle art of diagnostic testing revisited. In E. Shohamy, & A. Ronald Walton (Eds.), *Language assessment for feedback: testing and other strategies* (pp. 29-41). USA: Kendall/Hunt Publishing Company.

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5-15.

Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26, 233-246.

Sullivan, P. N. (2000). Spoken artistry: Performance in a foreign language classroom. In J. K. Hall, & L. S. Verplaetse (Eds.), *Second and foreign language learning through classroom interaction* (pp. 60-75). Mahwah, NJ: Lawrence Erlbaum Associates.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass, & C. Madden (Eds.), *Input and second language acquisition* (pp. 91-103). Rowley, MA: Newbury House.

Swain, M. (1988). Manipulating and complementing content teaching to maximise second language learning. *TESL Canada Journal*, 6(1), 68-83

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review*, 50, 158-164.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds.), *Principle and Practice in Applied Linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Swain, M. (1996). Integrating language and content in immersion classrooms: Research perspectives. *The Canadian Modern Language Review*, 52(4), 529-548.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford: Oxford University Press.

Swain, M. (2001). Examining dialogue: another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, 18, 275-302.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook on Research in Second Language Learning and Teaching* (pp. 471-483). Mahwah NJ: Lawrence Erlbaum.

Swain, M., & Johnson, R.K. (1997). Immersion education: A category within bilingual education. In R.K. Johnson, & M. Swain (Eds.) *Immersion education: International perspectives* (pp. 1-16). NY: Cambridge University Press.

Swain, M., & Lapkin, S. (1982). *Evaluating bilingual education: A Canadian case study*. Clevedon, England: Multilingual Matters.

Swain, M. and Lapkin, S. (1995) Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16, 371-391.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16, 371- 391.

Taasoobshirazi, G., Zuiker, S. J., Anderson, K. T., & Hickey, D. T. (2006). Enhancing inquiry, understanding, and achievement in an astronomy multimedia learning environment. *Journal of Science Education and Technology*, 15(5), 383-395.

Tanner, H., & Jones, S. (2003). Self-efficacy in mathematics and students' use of self- regulated learning strategies during assessment events. In N. A. Pateman, B. J. Dougherty, & J. T. Zilliox (Eds.), *Proceedings of the 27th conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th conference of PME-NA* (pp. 275-282). University of Hawai'i.

Taras, M. (2003) To feedback or not to feedback in student self-assessment, Assessment and Evaluation in Higher Education, 28 (5), 549-565.

Taras, M. (2005). Assessment — summative and formative: Some theoretical reflections. *British Journal of Educational Studies*, 53, 466-478.

Taras, M. (2008). Issues of power and equity in two models of self-assessment. *Teaching in Higher Education*, 13(1), 81-92.

Tarone, E., & Swain, M. (1995). A sociolinguistic perspective on second language use in immersion classrooms. *The Modern Language Journal*, 79, 24-46.

Tharpe, R., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. New York: Cambridge University Press.

Torrance, H., & Pryor, J. (1995) Investigating teacher assessment in infant classrooms: methodological problems and emerging issues. *Assessment in Education*, 2(3), 305-320.

Torrance, H. & Pryor, J. (1998) *Investigating Formative Assessment: teaching, learning and assessment in the classroom.* Philadelphia, PA Open University Press.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615-631.

Tunsall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, 22(4), 389-404.

Van de Craen, P., Ceuleers, E., & Mondt, K. (2007). Cognitive development and bilingualism in primary schools: Teaching maths in a CLIL environment. In D. Marsh, & D. Wolff (Eds.), *Diverse contexts - converging goals. CLIL in Europe* (pp. 185-200). Frankfurt: Peter Lang.

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in Teacher-Student Interaction: A Decade of Research. *Educational Psychology Review*, 22, 271-296.

Van Lier, L. (1988). *The classroom and the language learner: ethnography and second language classroom research*. London: Longman.

Van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy and authenticity*. New York: Longman.

Van Lier, L. (2000). Constraints and resources in classroom talk: Issues in equality and symmetry. In C. N. Candlin, & N. Mercer (Eds.), *English language teaching in its social context: A reader* (pp. 90-107). New York: Routledge.

Veel, R., & Coffin, C. (1996). Learning to think like an historian: The language of secondary school history. In R. Hasan, & G. Williams (Eds.), *Literacy in society* (pp. 191-231). London: Longman.

Vollmer, H. J., Heine, L., Troschke, R., Coetzee, D., & Küttel, V. (2006). *Subject-specific competence and language use of CLIL learners: The case of geography in grade 10 of secondary schools in Germany*. Paper presented at the ESSE8 Conference in London, 29 August 2006.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Walsh, S. (2006). *Investigating classroom discourse*. New York: Routledge.

Walsh, S. (2012). Conceptualising classroom interactional competence. *Novitas-ROYAL*, 6(1), 1-14.

Walsh, S., & Jenks, C. (2010). An introduction to conceptualising learning in applied linguistics. In P. Seedhouse, S. Walsh, & C. Jenks (Eds.), *Conceptualising learning in applied linguistics* (pp. 1-5). Basingstoke: Palgrave Macmillan.

Walsh, S., & O'Keeffe, A. 2010. Investigating higher education seminar talk. *Novitas-ROYAL*, 4(2), 141-158.

Weeden, P., Winter, J., & Broadfoot, P. (2002) *Assessment: what's in it for schools?* London: Routledge Palmer.

Wells, G. (1985). *Language development in the pre-school years*. Cambridge: Cambridge University Press.

Wells, G. (1993). Re-evaluating the IRF sequence: A proposal for the articulation of theories of activity and discourse for the analysis of learning and teaching in the classroom. *Linguistics and Education*, 5(1), 1-37.

Wells, G. (1999). *Dialogic inquiry: towards a sociocultural practice and theory of education.* Cambridge: Cambridge University Press.

Wells, G. 2007. The mediating role of discoursing in activity. *Mind, culture, and Activity*, 14, 160-177.

Wenger, E. (1998). *Communities of practice*. Cambridge: Cambridge University Press.

Wertsch, J. V. (1979). From social interaction to higher psychological process: A clarification and application of Vygotsky's theory. *Human Development*, 22, 1-22.

White, B. Y., & Fredenckson, J. R. (2000). Metacognitive facilitation: An approach to making scientific inquiry accessible to all. In J. Minstrel, & E. van Zee (Eds.), *Inquiring into inquiry leaming and teaching in science* (pp. 331-370). Washington, DC: American Association for the Advancement of Science.

Whittaker, R., & Llinares, A. (2009). CLIL in Social Science Classrooms: Analysis of Written and Spoken Productions. In Y. Ruiz de Zarobe, & R. M. Jiménez Catalán (Eds.), *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 215-234). Bristol: Multilingual Matters.

Whittaker, R., Llinares, A., & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research*, 15(3), 343-362.

Widdowson, H. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

Wiliam, D. (2000b). Integrating summative and formative functions of assessment, Keynote address to the European Association for Educational Assessment, November, Prague, Czech Republic. Available online at: http://www.kcl.ac.uk//depsta/education/ hpages/dwliam.html

Wiliam, D. (2006). Formative assessment: getting the focus right. *Educational Assessment*, 11(3/4), 283-289.

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment. *British Educational Research Journal*, 22, 537-548.

Wiliam, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, 11, 49-65.

Wilkinson, S. S. (1981). The relationship of teacher praise and student achievement: A meta-analysis of selected research. *Dissertation Abstracts International*, 41(9-A), 3998.

Winne, P. H. 1979. Experiments relating teachers' use of higher cognitive questions to student achievement. *Review of Educational Research*, 49(1), 13-50.

Wolff, D. (2007). "CLIL: bridging the gap between school and working life". In D. Marsh, & D. Wolff (Eds), *Diverse contexts - converging goals. CLIL in Europe* (pp. 15-25). Frankfurt: Peter Lang.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17, 89-100.

Wood, D. (1992). Teaching talk. In Norman, K (Ed.), *Thinking voices: The work of the National Oracy Project* (pp. 203-14). London: Hodder & Stoughton.

Wragg, E. C. (1993). *Classroom Teaching Skills*. London: Routledge.

Wragg, E. C., & Brown, G. A. (2001). *Questioning in the primary school*. London: Routledge.

Young, R. (1992). *Critical Theory and Classroom Talk.* Clevedon, UK: Multilingual Matters Ltd.

Young, R. (1995). Conversational styles in language proficiency interviews. *Language Learning*, 45(1), 3-42.

Young, R., & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14, 403-424.

Zhang Waring, H. (2008). Using explicit positive assessment in the language classroom: IRF, feedback and learning opportunities. *The Modern Language Journal*, 92(4), 577-594.